Verallgemeinerte lineare Modelle in der empirischen Sozialforschung: NONMET/ GLIM Workshop, 16.-20.11.81

Küchler, Manfred (Ed.)

Veröffentlichungsversion / Published Version
Konferenzband / conference proceedings

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

VERALLGEMEINERTE LINEARE MODELLE
IN DER EMPIRISCHEN SOZIALFORSCHUNG

NONMET/GLIM Workshop

16.-20.11.81

ZUMA-Arbeitsbericht No. 1982/03, März 1982

Inhaltsverzeichnis

E I N L E I T U N G

Sozialwissenschaftliche Daten sind in ihrer großen Mehrheit nicht-metrischer Art, d.h. sie stellen keine Meßpunkte auf exakt definierten Skalen dar, sondern drücken i.a. lediglich qualitative Unterschiede aus.

Die formale Anwendung klassischer Analysetechniken (multiple Regression, Varianzanalyse, Faktorenanalyse, etc.) auf solche Daten ist äußerst problematisch, die Gefahr von reinen Methoden-Artefakten hier besonders groß. Glücklicherweise ist neben die unbefriedigende Alternative, auf komplexe Techniken ganz zu verzichten und sich mit zweidimensionalen Kreuztabellierungen zu bescheiden, seit einigen Jahren eine weitere getreten: Die Statistiker haben Verfahren entwickelt, die den klassischen Techniken in Elaboriertheit nicht nachstehen, gleichwohl aber mit realistischen Annahmen hinsichtlich des Meßniveaus auskommen.

In der Soziologie sind besonders die Arbeiten von Leo A. GOODMAN bekanntgeworden, der die statistischen Überlegungen zu log-linearen Modellen und damit verbundenen Maximum-Likelihood-Schätzungen mehrdimensionaler Häufigkeitsvektoren (Kreuztabellen) auch selbst auf sozialwissenschaftliche Daten und Problemstellungen angewandt und nicht zuletzt ein entsprechendes Computer-Programm (ECTA) schon in den frühen 70er Jahren allgemein zugänglich gemacht hat. Zum "Goodman-Ansatz" haben eine Reihe von Statistikern wertvolle Beiträge geliefert (insbesondere Bishop, Fienberg, Holland sowie Haberman); ein bequemes EDV-Programm (P3F) ist in das BMDP-Paket aufgenommen worden.

Parallel ist von GRIZZLE, STARMER und insbesondere Gary KOCH der nach den Initialen der Autoren benannte GSK-Ansatz entwickelt worden. Sein augenfälligster Vorzug gegenüber dem Goodman-Ansatz besteht in der Möglichkeit, auch direkt mit (teilgruppenspezifischen) Anteilswerten zu rechnen, also eine logarithmische Transformation ganz zu vermeiden. Ein benutzerfreundliches Programm (NONMET von Herbert KRITZER) steht zur Verfügung.

Die Vor- und Nachteile beider Ansätze sind in der deutschen Rezeption recht kontrovers diskutiert worden (u.a. verschiedene Beiträge von Langeheine einerseits und Küchler andererseits in der Zeitschrift für Soziologie). Rückblickend läßt sich feststellen, daß dabei vorschnell Unterschiede proklamiert worden sind, die nach dem jüngsten Forschungsstand nicht wirklich ein Differenzierungskriterium sind. So lassen sich auch mit dem GSK-Ansatz "Pfadanalysen" nicht-metrischen Typs rechnen (vgl. das Küchler/Wides Papier in diesem Bericht) bzw. lassen sich mit dem Goodman-Ansatz auch nicht-hierarchische Designs und konditionale Effekte berechnen. Während also schon die jeweils isolierte weitere Beschäftigung mit beiden Ansätzen zu mehr Gemeinsamkeiten - Konvergenz, wenn man so will - führt, ist in der sozialwissenschaftlichen Diskussion lange Zeit ein formalstatistischer Ansatz wenig beachtet worden, der zumindest in weiten Teilen einen gemeinsamen Rahmen für Goodman- wie GSK-Ansatz abgibt, und so eine formale Erklärung für die Konvergenz in der Forschungspraxis bietet.

Dieser Ansatz ist von NELDER und WEDDERBURN in Großbritannien entwickelt und in das GLIM (= General Linear Model) Programm umgesetzt worden.

Konzept dieser ZUMA-Arbeitstagung war es, diesen Ansatz in den Sozialwissenschaften bekanntzumachen, obwohl die große Allgemeinheit in formalstatistischer Hinsicht einer schnellen Einarbeitung nicht gerade förderlich ist. Innerhalb der einwöchigen Arbeitstagung waren deshalb 5 Doppelstunden einer systematischen Einführung in den GLIM-Ansatz gewidmet. Für diese Vorlesungen konnte Prof. Gerhard Arminger (Wuppertal) gewonnen werden. Eine nachträgliche Ausarbeitung der Vorlesungen ist in diesem Arbeitsbericht abgedruckt.

Da der GSK-Ansatz Gegenstand des Frühjahrsseminars '81 des Kölner Zentralarchivs war, sollte er im Rahmen dieser Arbeitstagung nur kurz exemplarisch vorgestellt werden, dafür aber verstärkt auf komplexere Anwendungen eingegangen und für interessierte Teilnehmer Gelegenheit geboten worden, eigene Analysen im sogenannten NONMET-Forum zur Diskussion zu stellen. Neben bereits ausgearbeiteten Analysen, die Heiner Meulemann (ZA Köln) und Erhard Schwedler (Uni Frankfurt) vorstellten, wurden eine Reihe von Detailproblemen erörtert, die sich bei der praktischen Arbeit verschiedener Teilnehmer ergeben hatten.

Schließlich, dies der dritte Aspekt in der Konzeption dieser Arbeitstagung, sollte den Teilnehmern Gelegenheit geboten werden, während des Workshops selbst praktisch mit NONMET und GLIM zu rechnen. Die ZUMA-Computerabteilung (Carol Cassidy) führte die Teilnehmer in die Benutzung der Mannheimer SIEMENS-Anlage ein und stand für technische Beratungen während der gesamten Tagung zur Verfügung.

In diesem Arbeitsbericht sind mehrere Papiere abgedruckt:

1) Die schon erwähnte Vorlesungsausarbeitung von Gerhard Arminger; im wesentlichen eine formalstatistische Darlegung des GLIM-Ansatzes auch im Vergleich zum Goodman- bzw. GSK-Ansatz.

2) Drei ergänzende Papiere zum GLIM-Ansatz von Horst Busse (BGA Berlin), die insbesondere die große Flexibilität durch die Einbindung benutzereigener Unterprogramme (MACROs) dokumentieren.
Für programmtechnisch weniger Versierte dürfte insbesondere die kurzgefaßte deutschsprachige Beschreibung der GLIM-Direktiven (Parameterkarten) interessant sein.

3) Ein für diesen Zweck um programmtechnische Exkurse erweitertes Papier von Küchler/Wides, das zum einen einen exemplarischen Einstieg in den GSK-Ansatz bietet und darüberhinaus die Möglichkeit einer nicht-metrischen "Pfadanalyse" innerhalb dieses Ansatzes demonstriert.

Da sich in allen Papieren ausführliche Literaturhinweise finden, soll in dieser Einleitung darauf verzichtet werden. Erwähnt werden aber sollen die Bezugsquellen für die Programme:

GLIM liegt gegenwärtig in Version 3 vor. Eine Version 4 ist für den Sommer 1982 angekündigt, die wesentliche Verbesserungen (Matrizenoperation) für das Erstellen verschiedener Unterprogramme (MACROS) bringen soll.
Nähere Auskünfte durch

Numerical Algorithms Group
7 Banbury Road
Oxford OX2 6NN
England

NONMET liegt z.Zt. in Version 6.12 vor. Größere Revisionen sind derzeit nicht geplant.
Alternative Programmumsetzungen sind GENCAT, auf das im sehr empfehlenswerten Lehrbuch von Forthofer und Lehnen (1981) ausführlich Bezug genommen wird sowie die Prozedur FUNCAT im SAS-System (nur IBM-Anlagen).

NONMET und GLIM sind von ZUMA für SIEMENS-Anlagen umgestellt worden. Nähere Auskünfte hierzu erteilt die ZUMA-Computerabteilung (Carol Cassidy).

Will man ein kurzes inhaltliches Fazit aus dem Verlauf der Arbeitstagung ziehen, so läßt sich feststellen, daß NONMET und GLIM eigentlich keine konkurrierende Ansätze repräsentieren, sondern der Einsatz des einen oder anderen Programms primär von der statistischen und EDV-mäßigen Vorbildung des potentiellen Benutzers abhängig gemacht werden sollte.

Beide Programme bieten die Möglichkeit mehrdimensionale Kreuztabellen wahlweise linear (mit Anteilswerten) oder log-linear zu analysieren. Im log-linearen Fall ist das Schätzverfahren von GLIM unter formalstatistischen Kriterien als überlegen anzusehen. Auswirkungen auf die substanzwissenschaftlichen Interpretationen sind jedoch nicht zu erwarten.

Beide Programme bieten durch geeignete Definition der sogenannten Design-Matrizen die Möglichkeit gemischte Analysen durchzuführen, also den metrischen Charakter einzelner Merkmale mitauszunutzen.

GLIM bietet gleichzeitig alle Möglichkeiten klassischer multivariater Analyse, die sich unter das Lineare Modell subsumieren lassen (Regression, Varianzanalyse, etc.). Der Benutzer kann es durch eigene Unterprogramme sehr flexibel nach seinen Bedürfnissen ausgestalten; dafür läßt der Standard-Output aber einige Wünsche offen. Benutzer mit guten statistischen Kenntnissen und etwas Programmiererfahrung werden die mit GLIM gebotenen Möglichkeiten am besten ausschöpfen können.

NONMET bietet keine metrischen Techniken; es ist als Spezialprogramm konzipiert, wobei dem Benutzerkomfort großes Gewicht beigelegt worden ist. In vielen Fällen können die Design-Matrizen durch einfache Angaben impliziert werden. Seine Handhabung stellt weniger hohe Ansprüche an statistische/mathematische Kenntnisse. Eigene Programmierergänzungen sind nicht vorgesehen und in Anbetracht eines sehr unübersichtlichen Quellencodes auch nahezu unmöglich.

Klassische Anwendungen verallgemeinerter linearer Modelle in der empirischen Sozialforschung

Prof. Dr. Gerhard Arminger

FB 6

Universität - Gesamthochschule
Wuppertal

Februar 1982

Inhaltsverzeichnis Seite

1. Einführung

1.1. Umriß des Problemes

"Das Gewebe dieser Welt ist aus Zufall und Notwendigkeit gebildet" (Goethe (1796)).

In der vergangenen Dekade wurden der empirischen Sozialforschung neben den klassischen statistischen Methoden etwa der Regressions-, Varianz- und Faktorenanalyse auch statistische Modelle zur Behandlung von Nominaldaten zur Verfügung gestellt. Der Ansatz der loglinearen Modelle, entwickelt von Birch (1963), Goodman (1970, 1972, 1978), Haberman (1974) und anderen sowie die von Grizzle, Starmer, Koch (1969) auf Nominaldaten angewandte Regression wurden auch im deutschen Sprachraum bald nach ihrer Entwicklung aufgegriffen (Arminger (1976, 1979), Küchler (1979), Langeheine (1980)) und in großen Projekten (z. B. Arminger, Lijphart, Müller(1981)) angewandt. Dabei haben sich vor allem die Programmpakete ECTA und NONMET durchgesetzt.

Wenig beachtet wurde hingegen die von Nelder und Wedderburn (1972) erfundene Verallgemeinerung der gebräuchlichen linearen Modelle (Regressions-, Varianz-, Kovarianzanalyse) mit dem dazugehörigen Programmpaket GLIM (Generalised linear interactive modelling(1978)), das neben der Regressions-, Varianz- und Kovarianzanalyse auch loglineare, logistische, probit und viele andere Modelle als Spezialfälle einschließt.

Damit können folgende Probleme gelöst werden, die in den bisher vorwiegend verwendeten Ansätzen Schwierigkeiten bereiten,

- die korrekte Behandlung fehlender Zellenwerte bei der Berechnung von Freiheitsgraden sowie von Haupt- und Interaktionseffekten in loglinearen Modellen. (Bekanntlich lassen sich diese Berechnungen in ECTA nur durchführen, wenn alle Zellen besetzt sind.)

- die Einbeziehung von quantitativen unabhängigen Variablen. (Die hier vorgeschlagene Einbeziehung kann auch auf NONMET übertragen werden.)

- bei loglinearen Modellen kann die geschätzte Varianz Kovarianzmatrix der Schätzer für Haupt- und Interaktionseffekte angegeben werden, so daß beliebige lineare Kontraste - z. B. Differenzen zwischen zwei Parametern - getestet werden können.

Weiters kann gezeigt werden, daß sich zumindest im Fall der in NONMET (1981) verwendeten Grundmodelle der GSK-Ansatz (Grizzle, Starmer, Koch (1969)) als Spezialfall verallgemeinerter linearer Modelle (in Zukunft als GLM Ansatz oder als GL Modelle bezeichnet) darstellen läßt, so daß sich insgesamt ein einheitliches überschaubares Modell für die meisten gebräuchlichen Analyseverfahren für quantitative und nominal skalierte Variable ergibt.

Wir beschränken uns im folgenden darauf, zu zeigen, daß die bisher gebräuchlichen Verfahren auf GL Modelle zurückgeführt werden können und daß die oben genannten Probleme mit dem GLM Ansatz gelöst werden können. Daher wurde die vorliegende Arbeit mit klassischen Anwendungen von GL Modellen betitelt, ein Ausblick auf weitere Modelle wird am Ende gegeben.

1.2. Ein einführendes Beispiel

Im Rahmen des VASMA Projektes (VASMA = Vergleichende Analyse der Sozialstruktur mit Massendaten) an der Universität Mannheim wird unter anderem die Entwicklung der Frauenerwerbstätigkeit im zeitlichen Verlauf untersucht. Wir entnehmen dem Datenbestand von VASMA[1)] folgende Tabelle, die die Erwerbstätigkeit von verheirateten Frauen der Geburtsjahrgänge 1931 - 1941 im Mikrozensus 1971 angibt. Die Ehemänner sind nicht selbständig. Um das Datenmaterial übersichtlich zu halten, wurden einige Ausprägungen weggelassen.

Die verwendeten Variablen und ihre Ausprägungen sind

A Schulbildung der Frau

- A1 Nur Volksschule
- A2 Mittlere Ausbildung
- A3 Höhere Ausbildung

B Kinder

- B1 Keine Kinder
- B2 Kinder, die älter als 6 Jahre sind
- B3 Kinder, die jünger als 6 Jahre sind

C Erwerbstätigkeit

- C1 Nicht erwerbstätig
- C2 Un/angelernte Arbeiterin
- C3 Ausführende Angestellte und Beamte

X Einkommen des Mannes (Monatsdurchschnitt)

- X1 2000 DM
- X2 1500 DM
- X3 1000 DM
- X4 700 DM
- X5 450 DM

Z Einkommen des Mannes als quantitative Variable aufgefaßt (in Hundert DM skaliert)

R1 Zahl der nicht erwerbstätigen Frauen in der Kombination ABX

N1 Gesamtzahl der Frauen in der Kombination ABX

---

1) Für die Überlassung der Daten sowie für zahlreiche anregende Gespräche bin ich Herrn Prof. Dr. Walter Müller, Universität Mannheim, zu großem Dank verpflichtet.

Die Tabelle nimmt dann folgende Gestalt an:

Tabelle 1.1: Erwerbstätigkeit verheirateter Frauen im Mikrozensus 1971

X	A	B	R1	N1	Z
5	1	1	0016	0032	04.5
5	1	2	0052	0096	04.5
5	1	3	0043	0057	04.5
5	2	1	0005	0016	04.5
5	2	2	0013	0035	04.5
5	2	3	0017	0026	04.5
4	1	1	0132	0383	07.0
4	1	2	0640	1155	07.0
4	1	3	0607	0793	07.0
4	2	1	0047	0217	07.0
4	2	2	0260	0461	07.0
4	2	3	0265	0364	07.0
4	3	1	0001	0003	07.0
4	3	3	0000	0001	07.0
3	1	1	0329	0845	10.0
3	1	2	2925	4398	10.0
3	1	3	2838	3359	10.0
3	2	1	0242	0913	10.0
3	2	2	1874	2926	10.0
3	2	3	2384	2877	10.0
3	3	1	0001	0013	10.0
3	3	2	0009	0014	10.0
3	3	3	0013	0015	10.0
2	1	1	0100	0207	15.0
2	1	2	0927	1246	15.0
2	1	3	1022	1126	15.0
2	2	1	0178	0617	15.0
2	2	2	1581	2036	15.0
2	2	3	2118	2420	15.0
2	3	1	0010	0023	15.0
2	3	2	0039	0051	20.0
2	3	3	0095	0109	15.0
1	1	1	0016	0032	20.0
1	1	2	0143	0162	20.0
1	1	3	0147	0153	20.0
1	2	1	0106	0199	20.0
1	2	2	0722	0820	20.0
1	2	3	0908	0960	20.0
1	3	1	0016	0023	20.0
1	3	2	0094	0102	20.0
1	3	3	0209	0217	20.0

Wir wollen nun die Zahl der nicht erwerbstätigen Frauen erklären durch die qualitativen unabhängigen Variablen Ausbildung (A) und Kind (B) und durch die quantitative Variable Einkommen des Ehemannes (Z).

Zu diesem Zweck verwenden wir ein logit Modell mit der abhängigen Variablen

$$\hat{\eta}_i = \ln \frac{Rl_i}{Nl_i - Rl_i}$$ für jede Kombination von ABZ, i=1,... 41.

und rechnen ein GLM, das analog zu einer gewichteten Kovarianzanalyse aufgebaut und interpretierbar ist.

Ist das aus einem GL Modell geschätzte $\hat{\eta}_i$ berechnet, erhalten wir einen Schätzwert für Rl aus der Umkehrtransformation

$$\hat{Rl}_i = Nl_i \hat{\pi}_i \quad \text{mit} \quad \hat{\pi}_i = e^{\hat{\eta}_i}/(1+e^{\hat{\eta}_i})$$

Ist $\eta_i = 0$, so ist $\pi_i = 0.5$

Zunächst untersuchen wir, in welcher Stärke die einzelnen Variablen und ihre Interaktionen die abhängige Variable beeinflussen und entscheiden damit die Frage eines geeigneten Modells zur Erklärung der Daten.

Tabelle 1.2.: Modelle zur Erklärung der Daten in Tabelle 1.1

Modell	Devianz	Freiheits-grade	Effekt der Variablen	Devianz gegenüber Basismodell BM	Freiheits-grade	erklärte Devianz B
BM1: A+B+Z+A.B+A.Z+B.Z+A.B.Z	60.46	23	-	-	-	-
BM2: A+B+Z+A.B+A.Z+A.Z	69.26	27	A.B.Z	BM1: 8.8	4	0.002
A+B+Z+A.Z+B.Z	118.9	31	A.B	BM2: 49.64	4	0.011
A+B+Z+A.B+B.Z	76.58	29	A.Z	BM2: 7.32	2	0.002
A+B+Z+A.B+A.Z	86.35	29	B.Z	BM2: 17.09	2	0.004
BM3: A+B+Z	167.0	35	A.B,A.Z,B.Z,A.B.Z	BM1: 106.54	12	0.024
B+Z	199.3	37	A	BM3: 32.3	2	0.007
A+Z	3508.0	37	B	BM3: 3341.0	2	0.748
A+B	883.5	36	Z	BM3: 716.5	1	0.160
BM4: GM	4469.0	40	-	-	-	-

Unter der Rubrik Modelle sind jeweils die am Modell beteiligten Variablen angegeben, wobei A,B,Z die Variablen selbst und A.B, A.Z etc. die Interaktionen sind.

Als Basismodelle fungieren

BM1 - alle Interaktionen sind zugelassen
BM2 - alle Zweier Interaktionen sind zugelassen
BM3 - nur Haupteffekte sind zugelassen
BM4 - nur die Regressionskonstante GM ist zugelassen.

Die einzelnen Variablen bzw. ihre Interaktionen werden jeweils auf die kleinste, sie einschließenden Basismodelle bezogen.

Die Devianzen geben die Abweichungen der erwarteten von den beobachteten Häufigkeiten an. Die Devianzen sind - wenn die unter dem Modell erwarteten Häufigkeiten nur zufällig von den beobachteten Werten abweichen - χ^2 verteilt mit den jeweils angegebenen Freiheitsgraden.

Einzelne Interaktionen, Variable und Variablengruppen können in ihrer Einflußstärke überprüft werden, indem die Basismodelle ohne die jeweiligen Variablen gerechnet werden. Die Differenz der Devianzen ist - wenn das Basismodell zutrifft und die überprüfte Variable keinen Einfluß hat - wiederum χ^2 verteilt mit der neben der Differenz stehenden Zahl von Freiheitsgraden.

Werden die durch die einzelnen Variablen erzeugten Devianzen durch die Gesamtdevianz - die durch BM4 erzeugt wird - dividiert, erhalten wir den Anteil an erklärter Devianz, analog zum Anteil an erklärter Varianz.

Im obigen Beispiel liefert keines der Modelle eine ausreichende Anpassung an die Daten. Dies bedeutet, daß die quantitive Variable Z keinen linearen Zusammenhang mit den logits aufweist; wir werden das zeigen können, wenn wir das Einkommen als qualitative Variable auffassen.

Die einzelnen Variablen und Interaktionen haben alle signifikanten Einfluß, wenn wir eine Irrtumswahrscheinlichkeit von z. B. $\alpha = 0.05$ ansetzen und die Problematik nacheinander durchgeführter statistischer Tests im Moment außer acht lassen. Die Einflußstärke der einzelnen Variablen ist jedoch höchst unterschiedlich, wie die erklärten Devianzen zeigen. Den überragenden Einfluß nimmt die Variable B (Kind) mit $B_B = 0.748$, wesentlich geringer sind die Stärken von Z (Einkommen des Mannes) mit $B_Z = 0.16$ und A (Ausbildung) mit $B_A = 0.024$.

Um die Richtung des Einflusses der einzelnen Variablen zu sehen, berechnen wir in Modell A+B+Z+A.B+B.Z, das mit wenigen Variablen eine hohe Erklärungskraft besitzt, die Parameter mit Hilfe von GLIM.

Tabelle 1.3.: Schätzungen der Parameter, ihrer Standardabweichungen und der Standardabweichungen der Differenzen im Modell A+B+Z+A.B+A.Z aus Tabelle 1.2.

	ESTIMATE	S.E.	PARAMETER
1	1.208	.1179	<GM
2	-.4425	.7753E-01	A(2)
3	-.1923	.2699	A(3)
4	.4010	.1361	B(2)
5	1.640	.1500	B(3)
6	.7784E-01	.1040E-01	Z
7	.6016	.8706E-01	A(2).B(2)
8	.4443	.9443E-01	A(2).B(3)
9	.1969	.3529	A(3).B(2)
10	.1457	.3482	A(3).B(3)
11	.4653E-01	.1211E-01	B(2).Z
12	.4569E-01	.1334E-01	B(3).Z

S.E. OF DIFFERENCES

1	0.						
2	.1416	0.					
3	.2936	.2650	0.				
4	.2454	.1562	.3126	0.			
5	.2534	.1684	.3189	.1149	0.		
6	.1273	8.1639E-02	.2723	.1282	.1429	0.	
7	.1461	.1600	.2984	.1619	.1738	8.4512E-02	
8	.1506	.1642	.3007	.1661	.1787	9.2086E-02	
9	.3804	.3730	.5857	.3627	.3751	.3513	
10	.3761	.3686	.5829	.3653	.3614	.3466	
11	.1089	7.4920E-02	.2680	.1471	.1576	2.1707E-02	
12	.1090	7.5129E-02	.2681	.1445	.1621	2.2420E-02	
	1	2	3	4	5	6	
7	0.						
8	6.6879E-02	0.					
9	.3479	.3533	0.				
10	.3466	.3420	.3162	0.			
11	9.1665E-02	9.8022E-02	.3556	.3501	0.		
12	9.1117E-02	9.9424E-02	.3548	.3515	1.0397E-02	0.	
	7	8	9	10	11	12	

Die geschätzten Werte der einzelnen Parameter (Estimate), ihre Standardabweichungen (S.E. = standard error) und Bezeichnungen sind angegeben, die untere Dreiecksmatrix enthält die Standardabweichungen aller Paare von Differenzen.

Analog zur Varianzanalyse sind die Schätzer für Haupt- und Interaktionseffekte der qualitativen unabhängigen Variablen reparametrisiert, in GLIM werden die Effekte für die jeweils ersten Ausprägungen, also für A1, B1 und ihre Interaktionen 0 gesetzt.

Die Interpretation ist daher folgende:

GM ist die Regressionskonstante. Sie gibt - nach Transformation - die Wahrscheinlichkeit an, daß eine Frau mit Volksschulbildung (A1), ohne Kind (B1) und keinem Einkommen des Ehemannes (Z = 0) nicht erwerbstätig ist.

$$\hat{\pi} = \exp(GM)/(1+\exp(GM)) = \exp(-1.208)/(1+\exp(-1.208)) = 0.229.$$

Im Vergleich zu A1 - nur Volksschulbildung - wird die Wahrscheinlichkeit, nicht erwerbstätig zu sein, verringert, wenn die Frau mittlere Ausbildung (A(2) = -0.6325) oder höhere Ausbildung (A3) = -0.1923) besitzt. Allerdings ist A(3) nicht signifikant von 0 verschieden.

Man kann nämlich davon ausgehen, daß die Parameter in guter Näherung normalverteilt sind mit der angegebenen Standardabweichung. Bildet man das 95 % Konfidenzintervall um -0.1923 erhalten wir

$$P(-0.1923 - 1.96 \cdot 0.2699 \leq A(3) \leq -0.1923 + 1.96 \cdot 0.2699) \geq 0.95$$

Da der Wert 0 in diesem Intervall enthalten ist, läßt sich die H_0: A(3) = 0 gegen H_1: A(3) $\neq$ 0 nicht mit Irrtumswahrscheinlichkeit α = 0.05 verwerfen.

Die Variable B beeinflußt die Erwerbstätigkeit wie folgt: Kinder über 6 (B(2) = .6010) und Kinder unter 6 erhöhen die Wahrscheinlichkeit, nicht erwerbstätig zu sein, beträchtlich.

Als Beispiel berechnen wir die Wahrscheinlichkeit der Nichterwerbstätigkeit, wenn eine Frau Volksschulausbildung hat, Kinder unter 6 im Haus sind und der Ehemann ein Einkommen von Z = 10 = 1000 DM hat.

$$\hat{\eta} = GM + A(1) + B(3) + b \cdot Z$$
$$= -1.208 + 0 + 1.640 + 0.07784 \cdot 10$$
$$= 1.2104$$

$$\hat{\pi} = \exp(1.2104)/(1+\exp(1.2104))$$
$$= 0.77$$

Der Parameter von Z = 0.07784 gibt den Einfluß des Einkommens an. Der Regressionskoeffizient ist positiv, d.h. je höher das Einkommen, umso höher ist die Wahrscheinlichkeit der Nichterwerbstätigkeit. Der Effekt von Z ist ebenfalls beträchtlich, da ja jeweils das Einkommen mit diesem Koeffizienten multipliziert werden muß.

Die Interaktionen erhöhen jeweils die Wahrscheinlichkeit der Nichterwerbstätigkeit. Die Kombination A2B2, also mittlere Ausbildung und Kinder über 6 erhöht die Wahrscheinlichkeit, nicht erwerbstätig zu sein im Ausmaß von A(2).B(2) = 0.6016. Die Effekte von A(3)·B(2) und A(3)·B(3) sind nicht signifikant von 0 verschieden.

Die Interaktionen B(2)·Z = 0.04653 und B(3)·Z = 0.04569 zeigen, daß der Regressionskoeffizient von Z erhöht wird, wenn Kinder im Haus sind. Der positive Effekt des Einkommens auf Nichterwerbstätigkeit der Frau wird in diesem Fall stärker.

Wir testen noch, ob sich B(2).Z und B(3).Z voneinander unterscheiden.

H_0: $\Delta = B(2).Z - B(3).Z = 0$ gegen H_1: $\Delta \neq 0$

Da $\bar{\Delta} = 0.04653 - 0.04569 = 0.00084$ und die Standardabweichung der Differenz zwischen Parameter 11 und 12 = 0.010397 beträgt, kann H_0 nicht mit z. B. $\alpha = 0.05$ abgelehnt werden.

Insgesamt ist die Interpretation dieser Werte völlig analog zur Kovarianzanalyse zu bewerkstelligen. Ungewohnt ist zunächst nur die logit Transformation, sie verhindert jedoch, daß z. B. bei sehr hohem Einkommen Schätzwerte für die Wahrscheinlichkeit der Nichterwerbstätigkeit auftreten können, die größer als 1 sind. Weitere Vorteile werden im nächsten Abschnitt erörtert.

Einen weiteren Begriff von der Güte der Anpassung erhalten wir, wenn wir die beobachteten Werte von R1 (observed) mit den unter dem Modell erwarteten Werten (fitted) von R1 vergleichen. Wie wir später zeigen werden, sind die Residuen standardisiert, d.h. sie sind - wenn das Modell zutrifft - annähernd N(0.1) verteilt. Als Faustregel bedeutet dies, daß Werte, die dem Betrag nach größer als 2 sind, nur ca. 5 % aller Residuen ausmachen sollen.

In unserem Beispiel ist diese Faustregel verletzt, wie bereits aus der mangelnden Anpassung zu erwarten war.

Tabelle 1.4.: Beobachtete und erwartete Werte und standardisierte Fehler im Modell A+B+Z+A.B.+B.Z

UNIT	OBSERVED	OUT OF	FITTED	RESIDUAL
1	16	32	9.529	2.502
2	52	96	46.86	1.050
3	43	57	41.54	.4363
4	5	16	2.942	1.328
5	13	35	16.81	-1.290
6	17	26	17.95	-.4027
7	132	383	130.2	.1919
8	640	1155	653.1	-.7788
9	607	793	622.7	-1.361
10	47	217	46.63	.6160E-01
11	260	461	257.2	.2654
12	265	364	273.8	-1.070
13	1	3	.8948	.1328
14	0	1	.7773	-1.868
15	329	845	333.1	-.2877
16	2925	4398	2876.	1.549
17	2838	3359	2826.	.5812
18	242	913	234.5	.5660
19	1874	2926	1893.	-.7317
20	2384	2877	2344.	1.917
21	1	13	4.541	-2.060
22	9	14	9.170	-.9564E-01
23	13	15	12.52	.3314
24	100	207	101.4	-.1950
25	927	1246	970.3	-2.957
26	1022	1126	1022.	.7683E-03
27	178	617	208.4	-2.590
28	1581	2036	1575.	.3398
29	2118	2420	2156.	-2.458
30	10	23	10.17	-.7013E-01
31	39	51	44.28	-2.184
32	95	109	98.50	-1.135
33	16	32	18.76	-.9912
34	143	162	140.6	.5662
35	147	153	145.0	.7132
36	106	199	85.47	2.940
37	722	820	708.5	1.374
38	908	960	900.5	1.007
39	16	23	12.40	1.507
40	94	102	88.55	1.594
41	209	217	205.2	1.137

Wir fassen nun das Einkommen des Mannes als qualitative Variable mit 5 Einkommensklassen auf und berechnen das Modell A+B+X+A.B+B.X für die logits, das dem obigen Modell entspricht, wenn X statt Z gesetzt wird. Wir erhalten als Devianz einen Wert von D = 31,51 bei 20 Freiheitsgraden, also ein Modell mit wesentlich höherer Anpassung.

Die Parameter und ihre Standardabweichungen sind wie vorhin angegeben.

Tabelle 1.5.: Parameter und Standardabweichungen im Modell A+B+X+A.B+B.X für die Daten aus Tabelle 1.1.

	ESTIMATE	S.E.	PARAMETER
1	.6873	.1452	<GM
2	-.6216	.7775E-01	A(2)
3	-.2557	.2749	A(3)
4	1.361	.1776	B(2)
5	2.415	.2008	B(3)
6	-.9009	.1476	X(2)
7	-1.122	.1417	X(3)
8	-1.335	.1627	X(4)
9	-.7381	.3238	X(5)
10	.5901	.8724E-01	A(2).B(2)
11	.4364	.9465E-01	A(2).B(3)
12	.4095	.3568	A(3).B(2)
13	.1643	.3537	A(3).B(3)
14	.4567E-01	.1811	B(2).X(2)
15	-.2717	.1742	B(2).X(3)
16	-.4751	.1971	B(2).X(4)
17	-1.317	.3811	B(2).X(5)
18	-.3220E-01	.2031	B(3).X(2)
19	-.2530	.1968	B(3).X(3)
20	-.5926	.2214	B(3).X(4)
21	-1.346	.4278	B(3).X(5)

GM gibt jetzt das logit für die Kombination A1, B1, und X1 = 2000 DM Einkommen an. Die Interpretation der Wirkung der unabhängigen Variablen ändert sich nicht. Die Parameter für das Einkommen zeigen deutlich die nichtlineare Wirkung des Einkommens. Gegenüber der höchsten Einkommensklasse X1 wird die Wahrscheinlichkeit der Nichterwerbstätigkeit in der untersten Einkommensklasse (X(5) = -0.7381) in schwächerem Ausmaß verringert als in den mittleren Einkommensklassen (X(3) = -1.122 und X(4) = -1.335).

Zum Vergleich geben wir auch noch die beobachteten und unter dem Modell erwarteten Werte sowie die standardisierten Residuen an.

Tabelle 1.6.: Beobachtete und erwartete Werte und standardisierte Fehler im Modell A+B+X+A.B.+B.X

UNIT	OBSERVED	OUT OF	FITTED	RESIDUAL
1	16	32	15.59	.1439
2	52	96	47.84	.8500
3	43	57	41.88	.3370
4	5	16	5.407	-.2151
5	13	35	17.16	-1.408
6	17	26	18.12	-.4794
7	132	383	131.5	.4945E-01
8	640	1155	645.8	-.3449
9	607	793	605.9	.9338E-01
10	47	217	47.59	-.9755E-01
11	260	461	254.2	.5450
12	265	364	265.4	-.4355E-01
13	1	3	.8649	.1722
14	0	1	.7472	-1.719
15	329	845	332.1	-.2176
16	2925	4398	2894.	.9895
17	2838	3359	2852.	-.6866
18	242	913	235.6	.4864
19	1874	2926	1904.	-1.181
20	2384	2877	2370.	.6753
21	1	13	4.341	-1.965
22	9	14	9.684	-.3961
23	13	15	12.56	.3105
24	100	207	92.49	1.051
25	927	1246	956.0	-1.943
26	1022	1126	1011.	1.123
27	178	617	186.7	-.7594
28	1581	2036	1551.	1.584
29	2118	2420	2128.	-.5955
30	10	23	8.850	.4929
31	39	51	40.47	-.5096
32	95	109	96.88	-.5718
33	16	32	21.29	-1.982
34	143	162	143.5	-.1213
35	147	153	146.4	.2304
36	106	199	102.8	.4589
37	722	820	723.7	-.1807
38	908	960	910.8	-.4035
39	16	23	13.94	.8775
40	94	102	91.84	.7133
41	209	217	206.8	.6999

2. Struktur der verallgemeinerten linearen Modelle

2.1. Modellformulierung

Bei der Darstellung folgen wir den Ausführungen von Nelder und Wedderburn (1972) und Nelder (1981).

Wir nehmen an, daß n voneinander statistisch unabhängige Beobachtungen y_i, $i = 1,2 \ldots n$ vorliegen, für die gilt:

i) y_i folgt einer Wahrscheinlichkeitsverteilung
mit Varianz $V(y_i) = \sigma_i^2$, $i = 1, \ldots n$
und Erwartungswert $E(y_i) = \mu_i$, $i = 1, \ldots n$, so daß y_i als

(2.1) $y_i = \mu_i + e_i$ mit $E(e_i) = 0$, $V(e_i) = \sigma_i^2$, $i = 1, \ldots n$
mit e_i als Fehlerkomponente geschrieben werden kann.

ii) Die unabhängigen Variablen $x_{i1}, x_{i2} \ldots x_{ip}$, die als Kovariante beliebige Werte oder als Dummy-Variable bestimmte Werte z. B. 0, 1 annehmen können, definieren mit den unbekannten Regressionskoeffizienten β_j, $j = 1, \ldots p$ einen linearen Prädiktor oder systematische Komponente

$$(2.2)\quad \eta_i = \sum_{j=1}^{p} \beta_j x_{ij} \qquad i = 1, \ldots n$$

iii) Systematische und Fehlerkomponente werden durch eine Verbindungsfunktion (link) g verknüpft.

$$(2.3)\quad \eta_i = g(\mu_i) \qquad i = 1, \ldots n$$

$$y_i = g^{-1}(\eta_i) + e_i \qquad i = 1, \ldots n$$

Wir setzen voraus:

g ist monoton und zweimal stetig differenzierbar
$p(y_i)$ - die Dichtefunktion von y_i - entstammt der exponentiellen Familie, d.h.

(2.4) $p(y) = \exp\{|y\theta - b(\theta)|/a(\phi) + c(y,\phi)\}$

für Funktionen $a(\phi)$, $b(\theta)$ und $c(y,\phi)$,
θ wird als kanonischer und ϕ als Skalenparameter bezeichnet.

Betrachten wir alle Beobachtungen y_i , $i = 1, \ldots n$ gemeinsam, so läßt sich das Modell in Matrixschreibweise darstellen.
Vektoren und Matrizen kennzeichnen wir durch Unterstreichen.

(2.1') $\underline{y} = \underline{\mu} + \underline{e}$ mit $\underline{y} = (y_i)_{i=1,\ldots n}$, $\underline{\mu} = (\mu_i)_{i=1,\ldots n}$, $\underline{e} = (e_i)_{i=1,\ldots n}$ als Spaltenvektoren

(2.2') $\underline{\eta} = \underline{X\beta}$ $\underline{\eta} = (\eta_i)_{i=1,\ldots n}$, $\underline{X} = (x_{ij})_{\substack{i=1,\ldots n \\ j=1,\ldots p}}$, $\underline{\beta} = (\beta_j)_{j=1,\ldots p}$

(2.3') $\eta_i = g(\mu_i)$ $i = 1, \ldots n$

$\underline{\eta}$ und $\underline{\beta}$ sind Spaltenvektoren

$\underset{n\times p}{\underline{X}}$ ist die Matrix der Werte der p unabhängigen Variablen, bei kontrollierten Experimenten wird sie als Designmatrix bezeichnet.

2.2. Eigenschaften und spezielle Verteilungen der exponentiellen Familie

Es läßt sich zeigen (Beweis im Anhang A1), daß allgemein für die exponentielle Familie gilt

(2.5) $E(y_i) = \mu_i = b'(\theta_i)$ mit $b'(\theta_i) = \frac{db(\theta_i)}{d\theta_i}$ an der Stelle θ_i

(2.6) $V(y_i) = V_i = b''(\theta_i)a_i(\phi)$

Die Varianz läßt sich daher als Produkt von Funktionen des kanonischen Parameters θ und des Skalenparameters ϕ - unabhängig von θ - schreiben. $b''(\theta)$ wird als Varianzfunktion bezeichnet. $a_i(\phi)$ läßt sich häufig in der Form $a_i(\phi) = \frac{\phi}{w_i}$, wobei w_i Gewichte sind, darstellen.

Die für unsere Zwecke wichtigsten - aber bei weitem nicht alle - Verteilungen, die Spezialfälle der exponentiellen Familie darstellen, sind in der folgenden Tabelle angegeben.

Tabelle 2.1: Spezielle Verteilungen der exponentiellen Familie

	Normal	Poisson	Binomial	Gamma
Wertebereich von y	$(-\infty,\infty)$	0,1,2...	(0,1)	$(0,\infty)$
$a(\phi)$	ϕ	1	$\frac{1}{m}$ [1)]	$-\phi$
$b(\theta)$	$\frac{1}{2}\theta^2$	e^θ	$\ln(1+e^\theta)$	$\ln\ \theta$
$c(y,\phi)$	$-\frac{1}{2}(\frac{y^2}{\phi}+\ln 2\pi\phi)$	$-\ln\lvert y\rvert$	$\ln\binom{m}{my}$	$(\phi-1)\ \ln(y\phi)$ $+\ \ln\phi\ -\ \ln\Gamma(\phi)$
$\mu = E(y)$	θ	e^θ	$e^\theta/(1+e^\theta)$	$\frac{1}{\theta}$
Varianzfunktion $b''(\theta)$	1	μ	$\mu(1-\mu)$	$-\mu^2$

Die Normalverteilung liegt den klassischen Regressionsmodellen zugrunde. In ihrem Fall ist $V(y_i) = \sigma^2 = \phi$.

[1)] m ist die Größe der Stichprobe der relativen Häufigkeit y mit $Ey = \mu$.

Die Gammaverteilung, die als Spezialfall die χ^2-Verteilung enthält, läßt sich z. B. für schiefe Fehlerverteilungen anwenden.

Die Binomialverteilung ist Grundlage vieler Modelle mit dichotomer abhängiger Variablen (z. B. logit, probit, komplementäre log log Modelle).

Die Poissonverteilung ist die Verteilung der loglinearen Modelle, da die Dichte eines multinomial verteilten Zufallsvektors $(X_1, X_2 \ldots X_k)$ mit $\sum_{l=1}^{k} X_l = N$ als Produkt von k Dichten unabhängiger poissonverteilter Zufallsvariablen geschrieben werden kann, gegeben die Summe $\sum_{l=1}^{k} x_l = N$ ist ebenfalls poissonverteilt.

Der Beweis dieser Behauptung wird im Anhang A2 gegeben.

2.3. Link Funktion

Durch die link Funktion wird der lineare Prädiktor mit dem Erwartungswert der Beobachtung y verknüpft.

Im Fall einer normalverteilten Zufallsvariablen wird als Funktion die Identitätsfunktion $\eta_i = \mu_i$ gewählt.

Wir erhalten damit das klassische Regressionsmodell für den univariaten Fall

(2.7) $\underline{y} = \underline{\mu} + \underline{e}$

(2.8) $\underline{\eta} = \underline{\mu}$, ebenso ist $\underline{\theta} = \underline{\eta} = \underline{\mu}$

(2.9) $\underline{\eta} = \underline{X}\beta$

mit $E\underline{e}\underline{e}' = \sigma^2\underline{I}$ und $\underline{y} \sim N(\underline{X}\beta, \sigma^2\underline{I})$

Betrachten wir als abhängige Variable relative oder absolute Häufigkeiten, ergeben sich bei Wahl derselben link Funktion wie im normalverteilten Fall folgende Probleme:

i) der Wertebereich von μ, der auf (0,1) bei relativen Häufigkeiten bzw. auf 0,1, ... m bei absoluten Häufigkeiten beschränkt ist, kann leicht überschritten werden. Dies kann an folgendem Beispiel illustriert werden.

Sei μ die Wahrscheinlichkeit, als Ehefrau nicht erwerbstätig zu sein und sei x das Einkommen des Ehemannes. Es gelte die Regressionsgleichung $\mu = \beta_0 + \beta_1 x$.

Bei positivem β_1 braucht x nur genügend groß werden, so daß die Obergrenze 1 für μ mit Sicherheit überschritten wird.

Diese Überschreitung des zulässigen Wertebereichs kann auch bei Dummyvariablen etwa bei Kontingenztabellen auftreten.

ii) Da die Varianz von relativen Häufigkeiten im Binomialmodell $\frac{\mu(1-\mu)}{m}$ und von absoluten Häufigkeiten im Poissonmodell μ beträgt (vgl. Tab. 2.1), ist die im Regressionsmodell übliche kleinste Quadrate-Schätzung auf Grund der Verletzung der Forderung nach Homoskedastizität nicht mehr zulässig, es muß gewichtete Regression verwendet werden.

iii) Die identische link Funktion ermöglicht bei Häufigkeiten in einer Kontingenztabelle (poissonverteilte y_i) nicht die gewohnte Interpretation statistischer Unabhängigkeit und ihrer Erweiterungen wie Unabhängigkeit höherdimensionaler Verteilungen bei gegebenen Randverteilungen, bedingter Gleichverteilung und ähnlicher Konzepte. Dies läßt sich wiederum an einfachem Beispiel illustrieren.

Wir betrachten die 2 x 2 Kontingenztabelle mit y_{ik} als beobachtete Häufigkeiten und $Ey_{ik} = \mu_{ik}$, $i,k = 1,2$.

y_{11}	y_{12}	$y_{1.}$
y_{21}	y_{22}	$y_{2.}$
$y_{.1}$	$y_{.2}$	$y_{..}$

Unter der Hypothese der statistischen Unabhängigkeit lassen sich bei gegebenen Randverteilungen die erwarteten Werte

μ_{ik} $i,k = 1,2$ bekanntlich schreiben mit

$$\mu_{ik} = \frac{y_{i.}y_{.k}}{y_{..}}$$

Dieses Modell ist multiplikativ in μ_{ik}, aber additiv in $\ln\mu_{ik}$

$$\ln\mu_{ik} = -\ln y.. + \ln y_{i.} + \ln y_{.k}$$

Es liegt daher nahe, im Falle von absoluten Häufigkeiten die link Funktion

(2.10) $\eta = \ln\mu$

zu verwenden.

Betrachten wir noch einmal das Beispiel der Vierfeldertabelle.

Liegt statistische Unabhängigkeit vor, so ist das Kreuzprodukt

$$\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} = 1$$

Wir setzen die systematische Komponente

$$\eta_{ik} = \ln\mu_{ik} = \lambda + \alpha_i + \beta_k + \alpha\beta_{ik} \qquad i,k = 1,2$$

für die link Funktion $\eta = \ln\mu$ an.

Um bei 4 Beobachtungen alle 9 Parameter λ, α_i, β_k, $\alpha\beta_{ik}$ schätzen zu können, müssen wir 5 Restriktionen einführen (Reparametrisierung).

Wir setzen $\alpha_1 = \beta_1 = \alpha\beta_{11} = \alpha\beta_{12} = \alpha\beta_{21} = 0$

und erhalten folgendes Gleichungssystem

$\underline{\eta} = \underline{X\beta}$ mit

$$\underline{\eta} = \begin{bmatrix} \ln\mu_{11} \\ \ln\mu_{12} \\ \ln\mu_{21} \\ \ln\mu_{22} \end{bmatrix} \qquad \underline{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad \underline{\beta} = \begin{bmatrix} \lambda \\ \alpha_2 \\ \beta_2 \\ \alpha\beta_{22} \end{bmatrix}$$

Wir betrachten nun den Ausdruck

$$\ln \frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} = \ln\mu_{11} + \ln\mu_{22} - \ln\mu_{12} \ln\mu_{21}$$

$$= \lambda + \lambda + \alpha_2 + \beta_2 + \alpha\beta_{22} - \lambda - \beta_2 - \lambda - \alpha_2$$

$$= \alpha\beta_{22}$$

$\frac{\mu_{11}\mu_{12}}{\mu_{12}\mu_{21}} = 1$ genau dann, wenn $\ln \frac{\mu_{11}\mu_{12}}{\mu_{12}\mu_{21}} = \alpha\beta_{22} = 0$ wird.

Der statistischen Unabhängigkeit in der Vierfeldertafel ist das Nullsetzen der Interaktionsparameter im loglinearen Modell äquivalent.

Weiter gilt bei Verwendung der link Funktion $\eta = \ln\mu$, daß $\eta = \theta$ also gleich dem kanonischen Parameter der Poissonverteilung ist.

$\mu = e^{\theta} \rightarrow \theta = \eta = \ln\mu$

Mit dieser link Funktion lassen sich binomialverteilte Beobachtungen als Spezialfall multinomial (d.h. für uns poissonverteilte) verteilter Zufallsvariablen unter einem loglinearen Modell auffassen.

Wiederum soll dies ein einfaches Beispiel erläutern. Wir betrachten die dreidimensionale Tabelle, wobei wir C als abhängige Variable betrachten wollen.

		C_1	C_2
A_1	B_1	y_{111}	y_{112}
	B_2	y_{121}	y_{122}
A_2	B_1	y_{211}	y_{212}
	B_2	y_{221}	y_{222}
			N

Wir spezifizieren folgendes loglineare Modell

$$\eta_{ijk} = \ln\mu_{ijk} = \lambda + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk}$$

mit den Reparametrisierungen

$$\alpha_1 = \beta_1 = \gamma_1 = \alpha\beta_{11} = \alpha\beta_{12} = \alpha\beta_{21} = \alpha\gamma_{11} = \alpha\gamma_{12} = \alpha\gamma_{21} = \beta\gamma_{11} = \beta\gamma_{12} = \beta\gamma_{21} = 0$$

für die poissonverteilte Zufallsvariable y_{ijk}.

An Stelle von y_{ijk} betrachten wir jetzt die binomialverteilte Zufallsvariable

$$y_{ij} = \frac{y_{ij1}}{N} \quad \text{mit Erwartungswert} \quad \mu_{ij} = \frac{\mu_{ij1}}{N}$$

Das entsprechende logit Modell ist zunächst

$$\eta_{ij} = \ln \frac{\mu_{ij}}{1-\mu_{ij}} = \ln \frac{\mu_{ij1}/N}{1-\mu_{ij1}/N} = \ln \frac{\mu_{ij1}}{\mu_{ij2}}$$

Wir setzen nun das gewählte loglineare Modell ein.

$$\eta_{ij} = \ln \frac{\mu_{ij1}}{\mu_{ij2}} = \ln\mu_{ij1} - \ln\mu_{ij2}$$

$$= \lambda + \alpha_i + \beta_j + \gamma_1 + \alpha\beta_{ij} + \alpha\gamma_{i1} + \beta\gamma_{j1}$$

$$- \lambda - \alpha_i - \beta_j - \gamma_2 - \alpha\beta_{ij} - \alpha\gamma_{i2} - \beta\gamma_{j2}$$

$$= \gamma_1 - \gamma_2 + \alpha\gamma_{i1} - \alpha\gamma_{i2} + \beta\gamma_{j1} - \beta\gamma_{j2}$$

$$= -\gamma_2 - \alpha\gamma_{i2} - \beta\gamma_{j2} = \lambda^* + \alpha_i^* + \beta_j^*$$

†

Reparametrisierungsbedingungen

Aus dem loglinearen Modell lassen sich die Parameter des äquivalenten logistischen Modells sofort angeben, wobei die Parameter des logistischen Modells immer den um eine Ordnung höheren Parametern des loglinearen Modells entsprechen.

Den absoluten Häufigkeiten in einer Kontingenztabelle, die als unabhängige poissonverteilte Zufallsvariable aufgefaßt werden können, wird daher im GLM Ansatz die link Funktion $\eta = \ln\mu$ - also die Klasse der loglinearen Modelle - zugeordnet.

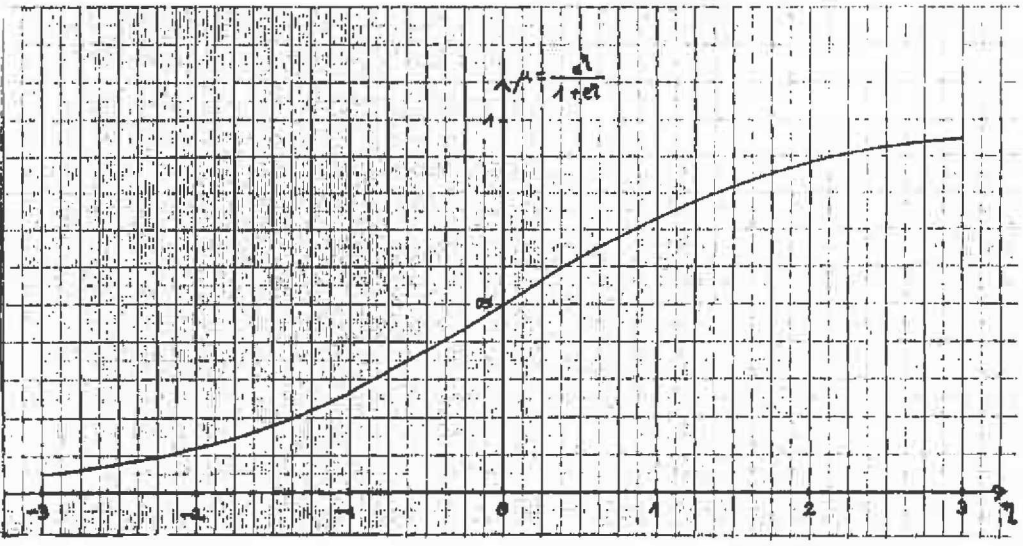
FUr binomialverteilte Zufallsvariable (relative Häufigkeiten y mit Erwartungswert μ) wurden mehrere link Funktionen entwickelt, von denen wir die wichtigsten angeben wollen.

(2.11) logit: $\eta = \ln \frac{\mu}{1-\mu}$

Da $\mu = \frac{e^{\theta}}{1+e^{\theta}}$ (vergl. Tab. 2.1) ist $\eta = \ln \frac{\mu}{1-\mu} = \theta$

wiederum der kanonische Parameter der Verteilung. Den Zusammenhang zwischen η und μ gibt Fig. 2.1 an.

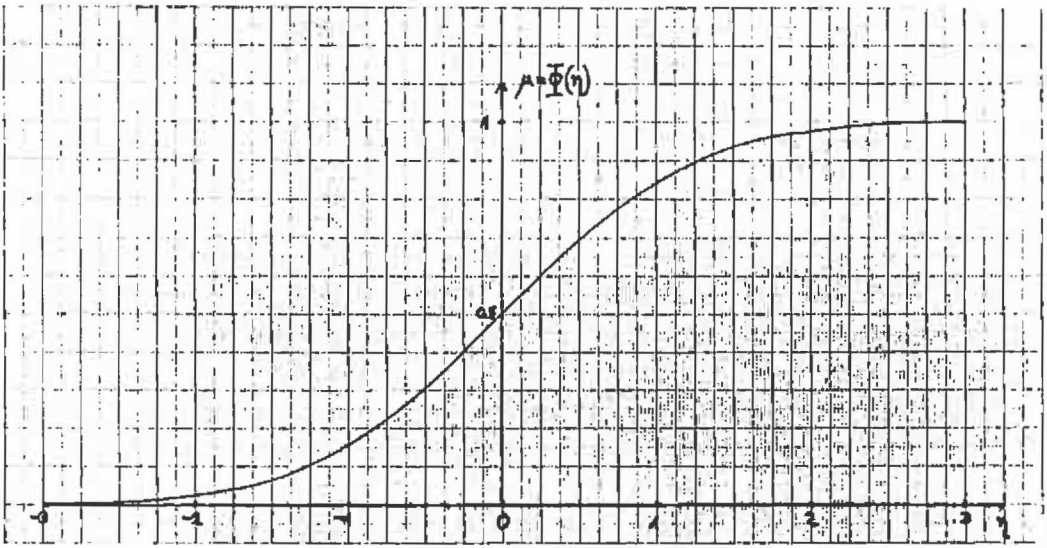
Fig. 2.1. Funktion logit



(2.12) probit: $\eta = \phi^{-1}(\mu)$

ϕ ist die Verteilungsfunktion der standardisierten Normalverteilung.

Fig. 2.2. Funktion probit



Logit und probit link Funktionen bilden den (0,1) Wertebereich von μ auf $(-\infty,\infty)$ ab, so daß keine Überschreitung des zulässigen Wertebereichs möglich ist und sind um den Wert 0 symmetrisch.

Beiden link Funktionen liegt die Annahme zu Grunde, daß es bei gleich bleibendem Ausmaß der Veränderung einer unabhängigen Variablen x_j an den Enden der Verteilung immer schwieriger wird noch Veränderungen der relativen Häufigkeit zu erzielen, da die Untergrenze 0 nicht unterschritten und die Obergrenze nicht überschritten werden darf (Bottom und Ceiling Effekte).

Komplementäre log log Funktion:

(2.13) $\eta = \ln(-\ln(1-\mu)) \quad \mu = 1-e^{-e^{\eta}}$

Diese link Funktion stammt aus folgender Überlegung der Epidemiologie. Die Wahrscheinlichkeit, daß eine Pflanze von einem Erreger infiziert wurde, sei sehr klein und betrage q. Bei n Erregern ist die Wahrscheinlichkeit, daß die Pflanze nicht infiziert wird, gleich

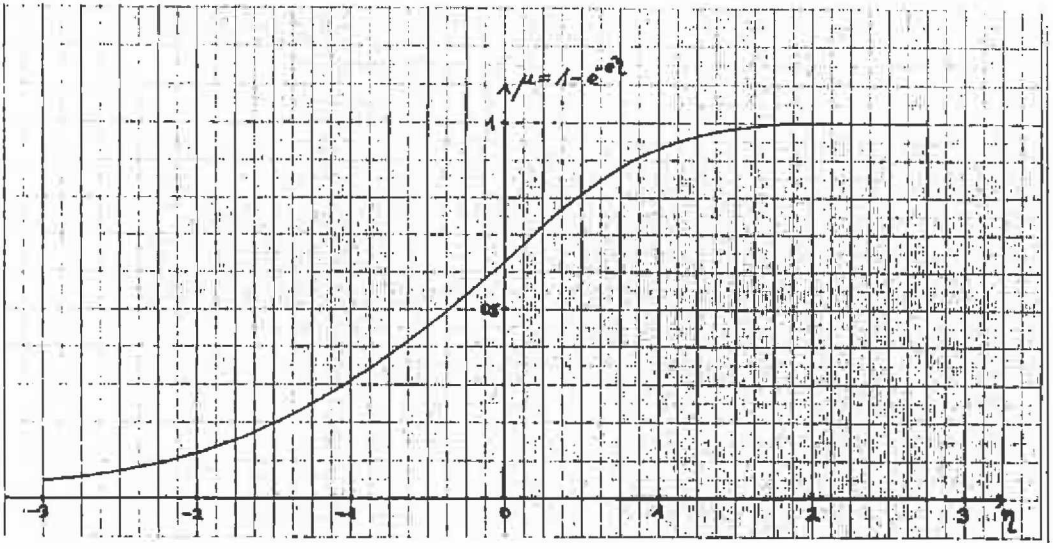
$(1-q)^n = (1-\frac{nq}{n})^n \sim e^{-nq}$.

Setzen wir $nq = \alpha$, so ist die Wahrscheinlichkeit μ einer Infektion gleich $1-e^{-\alpha}$. Da $\alpha > 0$ sein muß, setzen wir für α ein loglineares Modell mit Kovariaten an, so daß

$\eta = \ln\alpha = \sum_{j=1}^{p} \beta_j x_j \rightarrow \alpha = e^{\eta}$. Damit erhalten wir $\mu = 1-e^{-e^{\eta}}$

Diese Funktion ist asymmetrisch, aber für kleine Werte von von der logit Funktion nur wenig verschieden.

Fig. 2.3. Funktion komplementäres log log



Die link Funktion probit und komplementäre log log Funktion wurden für die Daten des einführenden Beispiels aus 1.2 verwendet. Die Ergebnisse sind in Anhang B ausgegeben. Die Güte der Anpassung sowie die standardisierten Fehler unterscheiden sich nur geringfügig von der link Funktion logit. Auf Grund der unterschiedlichen Transformation sind zwar die Schätzer für die Parameter verschieden, die inhaltliche Interpretation bleibt jedoch völlig gleich.

Für gammaverteilte Zufallsvariable wird gewöhnlich die link Funktion

(2.14) $\eta = 1/\mu$ verwendet, wobei wiederum $\eta = \theta$, also der kanonische Parameter der Gammaverteilung ist.

Im Programmsystem GLIM wird, soweit nichts anderes vorgesehen, jeweils die dem kanonischen Paramter entsprechende link Funktion zugeordnet, so daß folgende Verknüpfung gilt:

y-Variable	link Funktion
Normal	$\eta = \mu$
Poisson	$\eta = \ln\mu$
Binomial	$\eta = \ln(\mu/1-\mu)$
Gamma	$\eta = 1/\mu$

Für die uns speziell interessierenden Modelle poisson bzw. binomial verteilter Zufallsvariablen wurde erklärt, warum die Wahl der entsprechenden link Funktionen sinnvoll ist; trotzdem kann es sinnvoll sein, andere link Funktionen zu verwenden. Dies ist in GLIM auch vorgesehen, so daß der Benutzer in diesem Punkt sehr flexibel ist.

3. Schätzen und Testen im GLM-Ansatz

3.1. Berechnung der Regressionskoeffizienten

Ausgehend vom Modell

$$\underline{y} = \underline{\mu} + \underline{e}$$

$$\eta_i = g(\mu_i)$$

$$\underline{\eta} = \underline{X}\beta$$

stehen wir nun vor der Aufgabe, aus den uns bekannten $\underline{y}$, g und $\underline{X}$ den uns interessierenden Vektor der Regressionskoeffizienten $\underline{\beta}$ sowie die Abweichungen der beobachteten Werte von den unter dem Modell berechneten $\underline{\mu}$ Werten zu schätzen und ein Maß für die Güte der Anpassung des Modells an die beobachteten Daten zu finden. Wir bedienen uns dabei des Maximum Likelihood Schätzverfahrens und zur numerischen Berechnung der scoring Methode von Fisher. Dies führt zu folgenden Ergebnissen, die im mathematischen Anhang im einzelnen abgeleitet sind.

Die Berechnung des Regressionsvektors erfolgt in einem iterativen Verfahren der gewichteten Regression.

Sei $q = 0,1 \ldots$ der Laufindex der Iteration, so ist

$$(3.1)\quad \underline{b}^{q+1} = (\underline{X}'\underline{W}^q\underline{X})^{-1}\underline{X}'\underline{W}^q(\underline{\eta}^q+\underline{r}^q)$$

$\underline{b}^{q+1}$ ist der Schätzer von $\underline{\beta}$ in Iteration $q+1$

$$(3.2)\quad \underline{\eta}^q = \underline{X}\underline{b}^q$$

$$(3.3)\quad \underline{r}^q = (r_i^q)_{i=1,\ldots n} \text{ mit } r_i^q = (y_i-\mu_i^q)\frac{d\eta_i^q}{d\mu_i^q}$$

$$(3.4)\quad \mu_i^q = g^{-1}(\eta_i^q)$$

(3.5) $\frac{d\eta_i^q}{d\mu_i^q}$ = Ableitung von η_i nach μ_i an der Stelle μ_i^q. Diese ist natürlich von der gewählten link Funktion abhängig.

$\underline{W}^q$ = diag (w_i^q) $i = 1,\ldots n$ ist die Diagonalmatrix der Gewichte mit

(3.6) $w_i^q = \frac{1}{V(y_i)} \left(\frac{d\mu_i^q}{d\eta_i^q}\right)^2$ mit

$V(y_i)^q$ = Varianz von y_i, die sich aus der Varianzfunktion der Gewichte und dem Skalenparameter ergibt

$= b''(\theta_i)a_i(\phi)$

(3.7) $\frac{d\mu_i^q}{d\eta_i^q}$ = Ableitung von μ_i nach η_i an der Stelle η_i^q.

Die Iteration - bei der Normalverteilung ein Schritt, bei Binomial- und Poissonverteilungen in der Regel drei bis fünf Schritte - wird beendet, sobald

$$\sum_{j=1}^{p} |b_j^{q+1} - b_j^q| < \varepsilon$$

Die Anfangswerte werden gewonnen, indem die

$\mu_i^q = y_i \quad i = 1,\ldots n$

für $q = 0$ gesetzt werden.

Für den späteren Vergleich mit dem GSK Ansatz geben wir die Ableitungen für zwei spezielle Modelle an.

i) Normalverteilung

$$\eta_i = \mu_i,\ \mu_i^o = y_i,\ V(y_i) = \sigma^2,\ i = 1,2\ldots n$$

$$\frac{d\eta_i}{d\mu_i} = 1 \longrightarrow r_i^q = (y_i - \mu_i^q) \longrightarrow$$

$$(\eta^q + r^q) = (\mu_i^q + y_i - \mu_i^q) = y_i \quad \text{unabhängig von q}$$

$$w_i^q = \frac{1}{\sigma^2},\ \text{da}\ \left(\frac{d\mu_i^q}{d\eta_i^q}\right) = 1$$

Für die gewichtete Regression erhalten wir daher

$$\underline{b}^{q+1} = \sigma^2(\underline{X}'\underline{X})^{-1}\underline{X}'\frac{1}{\sigma^2}\underline{y} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y} \quad \text{unabhängig von q}$$

Das Verfahren ist daher nach dem ersten Schritt beendet, wir erhalten den gewohnten kleinsten Quadrate Schätzer als Ergebnis.

ii) Binomialverteilung, logit Funktion

$$\eta_i = \ln\frac{\mu_i}{1-\mu_i},\ \mu_i^o = y_i,\ V(y_i) = \frac{\mu_i(1-\mu_i)}{m_i},\ i = 1,2\ldots n$$

$$\frac{d\eta_i^q}{d\mu_i^q} = \frac{1}{\mu_i^q(1-\mu_i^q)}$$

$$\frac{d\mu_i^q}{d\eta_i^q} = \mu_i^q(1-\mu_i^q)$$

$$w_i^q = \left(\frac{\mu_i^q(1-\mu_i^q)}{m_i}\right)^{-1}(\mu_i^q(1-\mu_i^q))^2 = m_i\mu_i^q(1-\mu_i^q)$$

3.2 Konstruktion von Konfidenzintervallen und Tests der Regressionskoeffizienten

Auf Grund allgemeiner Ergebnisse der Maximum-Likelihoodschätzung gilt:

$\underline{b}$ ist asymptotisch normalverteilt mit Erwartungswert β und geschätzter Varianz Kovarianzmatrix $(\underline{X}'\underline{W}\underline{X})^{-1}$, wobei $\underline{b}$, $\underline{W}$ die Werte der letzten Interation sind. Dieses Resultat schreiben wir in Kurzform mit

$$(3.8)\quad \underline{b} \sim N(\underline{\beta}, (\underline{X}'\underline{W}\underline{X})^{-1}).$$

s_{jj} - die Varianz von b_j - ist dann der j-te Wert in der Diagonale von $(\underline{X}'\underline{W}\underline{X})^{-1}$.

Für b_j läßt sich nun (asymptotisch) sofort ein Konfidenzintervall bzw. ein Test angeben.

Beispiel: Zweiseitiges Konfidenzintervall zur Sicherheit $1-\alpha$.

$$P\left(b_j - z_{1-\frac{\alpha}{2}}\sqrt{s_{jj}} \leq \beta_j \leq b_j + z_{1-\frac{\alpha}{2}}\sqrt{s_{jj}}\right) = 1-\alpha \qquad z \sim N(0,1).$$

Beispiel: Test von H_o: $\beta_j = 0$ gegen H_1: $\beta_j \neq 0$

Ist $z^* = \frac{|b_j|}{\sqrt{s_{jj}}} \geq z_{1-\frac{\alpha}{2}}$, wird mit H_o mit Irrtumswahrscheinlichkeit α abgelehnt.

Um mehrere Parameter untereinander zu vergleichen, verwenden wir die Methode der linearen Kontraste.

Sei $\underline{c} = (c_j)$ $j = 1,\ldots p$ ein linearer Kontrast, d.h. $\sum_{j=1}^{p} c_j = 0$.

Für die Linearkombination $\underline{c}'\underline{b}$ gilt dann

(3.9) $\underline{c}'\underline{b} \sim N(\underline{c}'\underline{\beta}, \underline{c}'(\underline{X}'\underline{W}\underline{X})^{-1}c)$

Beispiel: $\underline{c}' = (1,-1,0 \ldots 0)$

$\underline{c}'\underline{b} = b_1 - b_2$. Damit kann sofort wie oben getestet werden, ob sich die beiden Parameter b_1, b_2 nur zufällig unterscheiden.

$\underline{c}' = (\frac{1}{2}, \frac{1}{2}, -1, 0 \ldots 0)$

$\underline{c}'\underline{b} = \frac{1}{2}b_1 + \frac{1}{2}b_2 - b_3$ ermöglicht den Test, ob sich der dritte Parameter vom arithmetischen Mittel der ersten beiden nur zufällig unterscheidet oder nicht.

Eine weitere Möglichkeit, zu testen, besteht darin, die Eigenschaft zu benutzen, daß unter der H_o: $\underline{c}'\underline{\beta} = 0$ der Ausdruck

(3.10) $(\underline{c}'\underline{b})^2/\underline{c}'\hat{\underline{\Sigma}}\underline{c}) = (\underline{c}'\underline{b}\underline{b}'\underline{c})/(\underline{c}'\hat{\underline{\Sigma}}\underline{c})$ mit $\hat{\underline{\Sigma}} = (\underline{X}'\underline{W}\underline{X})^{-1}$

asymptotisch χ_1^2 verteilt ist.

Diese Überlegung läßt sich für Gruppenvergleiche erweitern.

Sei $\underline{C}$ eine pxk Matrix mit Rang $\underline{C} = k \leq p$, so ist, wenn $\underline{b} \sim N(\underline{\beta}, \underline{\Sigma})$ verteilt ist,

(3.11) $\underline{C}'\underline{b} \sim N(\underline{C}'\underline{\beta}, \underline{C}'\hat{\underline{\Sigma}}\underline{C})$ ebenfalls multivariat normalverteilt.

Nach Anwendung eines Satzes von Anderson (1958, S. 54) ist dann der Ausdruck

(3.12) $(\underline{C}'\underline{b}-\underline{C}'\underline{\beta})'(\underline{C}'\hat{\underline{\Sigma}}\underline{C})^{-1}(\underline{C}'\underline{b}-\underline{C}'\underline{\beta}) \ldots \chi^2_k$ verteilt.

Beispiel: Wir wollen Überprüfen, ob in zwei Gruppen - Gruppe A durch Parameter b_1 und b_3, Gruppe B durch Parameter b_2 und b_4 gekennzeichnet - die Differenzen der Regressionskoeffizienten innerhalb der Gruppen jeweils gleich 0 sind.

Wir wählen

$$\underline{C}' = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \qquad \text{Rg } \underline{C} = 2$$

und setzen H_o: $\underline{C}'\underline{\beta} = \underline{0}$.

Der Ausdruck

$$(\underline{C}'\underline{b})'(\underline{C}'\hat{\underline{\Sigma}}\underline{C})^{-1}\underline{C}'\underline{b}$$

ist - wenn H_o zutrifft - χ^2 verteilt mit zwei Freiheitsgraden.

In der Varianzanalyse wird dies als Methode der multiplen Vergleiche bezeichnet (Scheffé, 1959).

Selbstverständlich gelten alle diese Eigenschaften nur asymptotisch, d.h. mit relativ großem Stichprobenumfang.

3.3. GÜte der Anpassung, Analoga zu multiplen und partiellen Bestimmtheitsmaßen

Unser Ziel ist es, die n Beobachtungen durch wenige unabhängige Variable x_{ij}, $j = 1,\ldots p$, also mit wenigen Parametern, zu erklären. Setzen wir alle erwarteten Werte $\hat{\mu}_i = \hat{\mu}$, so sprechen wir vom Minimalmodell (1 Parameter).

Setzen wir $\hat{\mu}_i = y_i$ so erhalten wir das saturierte oder volle Modell mit n Parametern, die sich sofort aus $\underset{n\times 1}{\underline{b}} = \underset{n\times n}{\underline{X}^{-1}}\underline{\eta}$ mit $\eta_i = g(y_i)$ berechnen lassen.

Das "richtige" Modell mit "hoher" Erklärungskraft und "wenigen" Parametern wird zwischen diesen Extremen liegen.

Da wir sowohl für das laufende Modell M_c mit p Parametern als auch das saturierte Modell M_f - indem wir ${}_f\mu_i = \hat{\mu}_i = y_i$ setzen - mit n Parametern Maximum-Likelihood-Schätzungen zur Verfügung haben, können wir die Theorie des Likelihood-Ratio-Tests heranziehen, um die GÜte der Anpassung des Modells an die Daten zu Überprüfen. Die Ableitungen des Testkriteriums allgemein und der Formeln für spezielle Verteilungen sind im Anhang A.4 enthalten. FÜr die Anwendung ist der folgende Satz von Bedeutung.

Trifft die Hypothese H_o: $\beta_{p+1} = \beta_{p+2} = \ldots = \beta_n = 0$ zu - unterscheiden sich also laufendes und saturiertes Modell nur zufällig - so ist die Teststatistik

$$(3.13)\quad S(c,f) = 2\sum_{i=1}^{n} \frac{1}{a_i(\phi)}\left[y_i({}_f\theta_i - {}_c\theta_i) + b({}_c\theta_i) - b({}_f\theta_i)\right]$$

χ^2 verteilt mit n-p Freiheitsgraden.

${}_f\theta_i$ und ${}_c\theta_i$ sind die kanonischen Parameter des saturierten bzw. laufenden Modells. S(c,f) heißt skalierte Devianz.

Ist $a_i(\phi) = \phi$ oder $\frac{\phi}{w_i}$, wobei w_i bekannte Gewichte sind, so wird

$$D(c,f) = \phi S(c,f)$$

als Devianz bezeichnet. D kann jeweils aus $\underline{b}$ berechnet werden und enthält keine Unbekannten mehr. Für die speziellen Verteilungen erhalten wir

i) Normalverteilung

$$a_i(\phi) = \sigma^2$$

$$(3.14)\quad S(c,f) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - {}_c\mu_i)^2$$

${}_c\mu_i$ sind die unter M_c geschätzten Werte von μ_i.

$$(3.15)\quad D(c,f) = \sum_{i=1}^{n} (y_i - {}_c\mu_i)^2$$

also die Summe der Abweichungsquadrate.

ii) Poissonverteilung

$$a_i(\phi) = 1$$

$$(3.16)\quad S(c,f) = 2 \sum_{i=1}^{n} \left(y_i \ln \frac{y_i}{{}_c\mu_i} + {}_c\mu_i - y_i\right)$$

Wegen der Maximierungsbedingung $\sum_{i=1}^{n} c^{\mu_i} = \sum_{i=1}^{n} y_i$ - die Randverteilungen der beobachteten und erwarteten Häufigkeiten in Kontingenztabellen müssen gleich sein (vgl. A.3) - erhalten wir das bekannte Testkriterium der loglinearen Modelle

$$G^2 = 2 \sum_{i=1}^{n} y_i \ln \frac{y_i}{c^{\mu_i}}$$ (vgl. Bishop et. al., 1975)

(3.17) $D(c,f) = S(c,f)$

iii) Binomialverteilung

$a_i(\phi) = \frac{1}{n_i} = \frac{\phi}{n_i}$, n_i = Stichprobengröße

$$(3.18)\quad S(c,f) = 2 \sum_{i=1}^{n} n_i \left[y_i \left(\ln \frac{y_i}{c^{\mu_i}} - \ln \frac{1-y_i}{1-c^{\mu_i}} \right) + \ln \frac{1}{1-c^{\mu_i}} - \ln \frac{1}{1-y_i} \right]$$

$$= 2 \sum_{i=1}^{n} \left[n_i y_i \ln \frac{y_i}{c^{\mu_i}} + (n_i - n_i y_i) \ln \frac{1-y_i}{1-c^{\mu_i}} \right]$$

(3.19) $D(c,f) = S(c,f)$

iv) Gammaverteilung

$a_i(\phi) = -\phi$

$$(3.20)\quad S(c,f) = \frac{1}{-\phi} 2 \sum_{i=1}^{n} \left(\ln \frac{y_i}{c^{\mu_i}} + \frac{y_i - c^{\mu_i}}{c^{\mu_i}} \right)$$

Wegen der Maximierungsbedingung (vgl. A.3) fällt üblicherweise der zweite Summand weg.

$$(3.21)\quad D(c,f) = 2 \sum_{i=1}^{n} \left(-\ln \frac{y_i}{c^{\mu_i}} + \frac{y_i - c^{\mu_i}}{c^{\mu_i}}\right)$$

Liegen zwei hierarchische geordnete laufende Modelle M_{c_1} und M_{c_2} vor; d.h. die Parameter $\beta_1 \ldots \beta_k$ von M_{c_1} sind eine Teilmenge der Parameter $\beta_1 \ldots \beta_p$ von M_{c_2}, so ist - wie im Anhang gezeigt wird -

$$(3.22)\quad S(c_1,c_2) = S(c_1,f) - S(c_2,f) \ldots \chi^2_{p-k} \text{ verteilt}$$

(asymptotisch) unter der Bedingung, daß H_o: $\beta_{k+1} \ldots \beta_n = 0$ ist.

Es lassen sich daher alle bei loglinearen oder linearen Modellen üblichen Strategien zum Auffinden gut passender Modelle auf verallgemeinerter lineare Modelle übertragen.

Bei der Verwendung normalverteilter Zufallsvariablen ist folgende Bemerkung geboten:

Der Wert von $\phi = \sigma^2$ ist in der Regel unbekannt.

Daher ist $S(c,f) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - c^{\mu_i})^2$ unbekannt.

Um trotzdem eine Hypothese H_1: $\beta_{k+1} = \beta_{k+2} \ldots \beta_p = 0$

testen zu können, bedient man sich der in der Varianzanalyse üblichen Überlegung.

Unter der Hypothese H_2: $\beta_{p+1} = \beta_{p+2} \ldots \beta_n = 0$ ist

$S(c_2,f)$ χ^2_{n-p} verteilt.

Treffen H_1 und H_2 zu, ist

$S(c_1,c_2) = S(c_1,f) - S(c_2,f)\ x^2_{p-k}$ verteilt.

Da $S(c_1,f) = \frac{1}{\sigma^2} D(c_1,f)$, $S(c_2,f) = \frac{1}{\sigma^2} D(c_2,f)$

$$S(c_1,c_2) = \frac{1}{\sigma^2} (D(c_1,f) - D(c_2,f)) = \frac{1}{\sigma^2} D(c_1,c_2)$$

ist der Ausdruck $\frac{S(c_1,c_2)/(p-k)}{S(c_2,f)/(n-p)} = \frac{D(c_1,c_2)/(p-k)}{D(c_2,f)/(n-p)}$

$F_{(p-k,n-p)}$ verteilt.

In D tritt jedoch σ^2 nicht mehr auf.

$D(c_1,c_2)$ ist die durch die Regression erklärte Quadratsumme, üblicherweise mit SSR bezeichnet.

$D(c_2,f)$ ist die Quadratsumme der Fehler, üblicherweise mit SSE bezeichnet.

Mit S bzw. D lassen sich zur Varianzanalyse analoge Tabellen aufstellen sowie multiple Bestimmtheitsmaße und PRE-Koeffizienten (vgl. Holm, 1979) bestimmen.

Sei M_{c_0} ein Basismodell, M_{c_1} und M_{c_2} laufende Modelle, die hierarchisch geordnet sind, so daß $M_{c_0} \subset M_{c_1} \subset M_{c_2}$.

$D(c_1,c_2) = D(c_1,f) - D(c_2,f)$ gibt die Devianz an, die durch die Parameter, die in M_{c_2}, aber nicht in M_{c_1} enthalten sind, erklärt wird. Ist M_{c_0} ein geeignetes Basismodell, so läßt analog zum multiplen Bestimmtheitsmaß

(3.23) $B = \frac{D(c_1,c_2)}{D(c_0,f)}$ als Anteil an erklärter Devianz

der Parametermenge $M_{c_2}-M_{c_1}$ auffassen.

Beispiel: Varianzanalyse

$D(c_0,f) = SST$

$D(c_1,c_2) = SSR$

$B = \frac{SSR}{SST} = R^2_{y\cdot x_1 \ldots x_p}$

Ein Analogon zum PRE-Koeffizienten läßt sich dann als multiples partielles Bestimmtheitsmaß konstruieren. (PRD = Prozentsatz an reduzierter Devianz.)

(3.24) $PRD = \frac{D(c_1,f)-D(c_2,f)}{D(c_1,f)} = \frac{D(c_1,c_2)}{D(c_1,f)}$

PRD gibt an, um welchen Prozentsatz die Devianz im M_{c_1} durch Einführung der zusätzlichen Variablen in M_{c_2} verringert wurde.

Im einführenden Beispiel haben wir Konfidenzintervalle, Tests und multiple Bestimmtheitsmaße sowohl für einzelne Variable als auch für Gruppen von Variablen angegeben.

3.4. Analyse der Residuen

Nach Berechnung der $\hat{\mu}_i$ können wir aus $y_i = \mu_i + e_i$ die Fehler $\hat{e}_i = y_i - \hat{\mu}_i$ schätzen. Zur Analyse dieser Residuen verwenden wir die sogenannten Pearson-Residuen, die wie folgt definiert sind:

$$(3.25)\quad {}_p e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(y_i)}} \qquad i = 1,2 \ldots n$$

Die Residuen werden durch ihre Standardabweichung dividiert.

Für die uns speziell interessierenden Verteilungen erhalten wir

i) Normalverteilung:

$$(3.26)\quad {}_p e_i = \frac{y_i - \hat{\mu}_i}{s} \quad \text{mit } s^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2$$

$$= \frac{1}{n-p} D(c,f)$$

wobei n-p die Freiheitsgrade des zur Berechnung von s^2 gewählten Modells M_c mit p unabhängigen Variablen sind.

ii) Poissonverteilung

$$(3.27)\quad {}_p e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$$, da $\hat{\mu}_i$ die geschätzte Varianz der poissonverteilten Zufallsvariablen ist.

Der Wert $\sum_{i=1}^{n} {}_p e_i^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} = \chi^2$ ist die übliche Restgröße des χ^2 Anpassungstests.

iii) Binomialverteilung

$$(3.28)\quad {}_{p}e_i = \frac{\sqrt{m_i}(y_i - \hat{\mu}_i)}{\sqrt{\hat{\mu}_i(1-\hat{\mu}_i)}}\,, \text{ da } \frac{\hat{\mu}_i(1-\hat{\mu}_i)}{m_i} \text{ die geschätzte Varianz}$$

der binomialverteilten Zufallsvariablen y_i ist.

Als Faustregel kann man nun annehmen, daß diese standardisierten Residuen in etwa normalverteilt sind. Abweichungen, die größer als $|\pm 2|$ sind, dürfen daher bei einem gut angepaßten Modell nur mit einer Wahrscheinlichkeit von ca. 5 % auftreten, so daß auf n = 100 nicht mehr als 5 große Abweichungen kommen sollen. Die standardisierten Fehler wurden auch in unserem Beispiel in Abschnitt 1.2 angegeben und analysiert.

Ein häufig auftretender Nachteil der Pearson-Residuen ist es, daß ihre Verteilung auch deutlich von der Normalverteilung verschieden sein kann. Es wurden daher verschiedene Versuche unternommen, Residuen zu definieren, die besser einer Normalverteilung folgen. Literaturhinweise zu diesem Punkt sind in Nelder (1981) enthalten.

4. Die Matrix der unabhängigen Variablen

Die Matrix $\underline{X}$, die die Werte der unabhängigen Variablen x_{ij}, $i = 1,...n$, $j = 1,...p$ enthält, bereitet Anwendern linearer Modelle erfahrungsgemäß große Schwierigkeiten. Wir werden daher an dieser Stelle versuchen, Grundprinzipien beim Aufbau von $\underline{X}$ zu erklären. (Vgl. Searle (1971), Bock (1975), Evers and Namboodiri (1979)).

Wir verwenden für die Analysen, die bestimmten Typen von $\underline{X}$ entsprechen, im folgenden die Bezeichnungen aus der Theorie linearer Modelle, also Regressions-Varianz, Kovarianzanalyse usw. Die Eigenschaften dieser Matrizen lassen sich durchgehend auf verallgemeinerte Modelle übertragen, an Stelle von Varianz- bzw. Kovarianzanalyse müßten wir dann von Devianz- bzw. Kodevianzanalyse sprechen.

Die Eigenschaften von $\underline{X}$ werden vor allem an kleinen Beispielen demonstriert, ausführliche formale Beschreibungen sind bei Searle (1971) oder Bock (1975) nachzulesen.

4.1. Regressionsanalyse

Ist

$$(4.1)\quad x_{ij} = \begin{cases} 1 \text{ für } j = 1,\ i = 1,...n \\ = \lambda_{ij} \text{ für } j = 2,...p,\ i = 1,... n \text{ mit } \lambda_{ij} \text{ beliebige reelle Zahl} \end{cases}$$

So haben wir die bekannte Regressionsanalyse vor uns.

In Gleichung (3.1) zur Berechnung von $\underline{b}$ tritt der Ausdruck $(\underline{X}'\underline{W}\underline{X})^{-1}$, also die Inverse einer pxp Matrix auf. Da $\underline{W}$ eine Diagonalmatrix mit $w_i \neq 0$, $i = 1,...n$ ist, ist $Rg\ \underline{W} = n$. Damit $(\underline{X}'\underline{W}\underline{X})^{-1}$ existiert, muß daher der Rang von $\underset{n\times p}{\underline{X}} = p$ sein, d.h. die Matrix $\underline{X}$ muß von vollem Spaltenrang sein. Ist dies nicht der Fall, ist die Determinante $|\underline{X}'\underline{W}\underline{X}| = 0$ und die Inverse ist nicht bestimmt.

Ist $|\underline{X}'\underline{W}\underline{X}|$ fast 0, sind die Schätzungen von $\underline{\beta}$ instabil. Dies tritt z. B. auf, wenn unabhängige Variable hoch untereinander korrelieren. Dieser Fall wird in der Ökonometrie unter dem Stichwort Multikollinearität ausführlich behandelt. (Vgl. etwa Judge et al., 1980).

Man bemerke, daß auch für poissonverteilte oder binomialverteilte Zufallsvariable die Matrix $\underline{X}$ ausschließlich quantitativ sein kann, die abhängigen Variablen sind in diesem Fall für jede Zeile von $\underline{X}$ absolute Häufigkeiten oder Prozentsätze.

Dies steht im Gegensatz zum üblichen zeilenweisen Einlesen eines Variablenvektors pro Person, wie es für quantitative abhängige Variable üblich ist.

4.2. Varianzanalyse

Sind die unabhängigen Variablen nominal skaliert, d.h. sie nehmen jeweils Merkmalsausprägungen A_j, B_k, C_l der Variablen A,B,C,... an, die nicht untereinander geordnet bzw. deren Abstände unbekannt sind, besitzt $\underline{X}$ eine besondere Gestalt.

4.2.1. Einfache Varianzanalyse

Betrachten wir zunächst nur eine unabhängige Variable A mit Ausprägungen $A_1, \ldots A_{q_1}, A_{q_1+1}$.

In unserem Modell gelte dann

$$(4.2)\quad \eta_i = \beta_0 + \beta_j \qquad j = 1, \ldots q_1+1,$$

je nachdem in welcher Kategorie von A_j sich i befindet.

Dies läßt sich auch in der Form schreiben

$$(4.3)\quad \eta_i = \beta_0 \cdot 1 + \beta_1 \cdot x_{i1} + \ldots + \beta_{q_1} x_{iq_1} + \beta_{q_1+1} x_{iq_1+1} \quad \text{mit}$$

$$(4.4)\quad x_{ij} = \begin{cases} 1 & \text{wenn } i \in A_j \\ 0 & \text{sonst.} \end{cases}$$

z. B. sei $n = 5$, $q_1+1 = 3$.

$$\begin{aligned} \eta_1 &= \beta_0 + \beta_1 \\ \eta_2 &= \beta_0 + \beta_1 \\ \eta_3 &= \beta_0 + \beta_2 \\ \eta_4 &= \beta_0 + \beta_3 \\ \eta_5 &= \beta_0 + \beta_3 \end{aligned} \quad \rightarrow \underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Über jeder Spalte von $\underline{X}$ stehen die zugehörigen Regressionskoeffizienten.

Wir sehen sofort, daß die erste Spalte die Summe der Spalten 2, 3, 4 von $\underline{X}$ ist. $\underline{X}$ erfüllt daher nicht die Bedingung, daß der Spaltenrang von $\underline{X}$ gleich der Anzahl der Spalten von $\underline{X}$ ist, die zur Berechnung der Inversen notwendig ist.

Damit $\underline{X}$ von vollem Spaltenrang ist, müssen wir lineare Restriktionen einführen, die auch als Reparametrisierungsbedingungen bezeichnet werden.

4.2.1.1. Zentrierte Effekte

Die am häufigsten verwendete lineare Restriktion der Parameter $\beta_1, \beta_2 \dots \beta_{q_1+1}$ ist die Forderung

$$(4.5) \quad \sum_{j=1}^{q_1+1} n_j \beta_j = 0 \rightarrow \beta_{q_1+1} = \frac{-1}{n_{q_1+1}} \sum_{j=1}^{q_1} n_j \beta_j$$

wobei n_j die Anzahl der Beobachtungen in der Kategorie A_j ist.

Setzen wir in unser obiges Beispiel ein, erhalten wir

$$\begin{aligned} n_1 &= \beta_0 + \beta_1 \\ n_2 &= \beta_0 + \beta_1 \\ n_3 &= \beta_0 + \beta_2 \\ n_4 &= \beta_0 - \beta_1 - \tfrac{1}{2}\beta_2 \\ n_5 &= \beta_0 - \beta_1 - \tfrac{1}{2}\beta_2 \end{aligned} \quad \rightarrow \underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_1 & \beta_2 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1/2 \\ 1 & -1 & -1/2 \end{bmatrix} \end{array}$$

mit Rg $\underline{X}$ = 3 also vollem Spaltenrang

In diesem Fall ist dann - wie durch Nachrechnen leicht gezeigt wird - b_o als Mittelwert der geschätzten Werte $\hat{n}_i$ und b_j als Abweichung der j-ten Kategorie vom Gesamtmittelwert interpretierbar. Die gewichtete Summe der Abweichungen ist dann 0.

In vielen Fällen, z. B. in Kontingenztabellen ist die Zahl der Beobachtungen pro Kategorie - d.h. die Anzahl der Zellen, nicht etwa der befragten Personen gleich, so daß die Forderung

$$\sum_{j=1}^{q_1+1} n_j \beta_j = 0 \quad \text{zu} \quad \sum_{j=1}^{q_1+1} \beta_j = 0$$

reduziert werden kann.

Beispiel: 2 x 3 Kontingenztabelle. y_{ij} sind die absoluten Häufigkeiten in der Zelle $A_i B_j$.

	B_1	B_2	B_3
A_1	y_{11}	y_{12}	y_{13}
A_2	y_{21}	y_{22}	y_{23}

Die Hypothese, daß die Beobachtungen von A unabhängig sind, läßt sich durch folgendes Modell darstellen:

$y_{ij} = \mu_{ij} + e_{ij}$ $\quad y_{ij}$ ist poissonverteilt

$n_{ij} = \ln \mu_{ij} = \beta_o + \beta_j$ mit $n_j = 2$, $\sum_{j=1}^{3} \beta_j = 0$.

$$
\begin{aligned}
\eta_{11} &= \beta_0 + \beta_1 \\
\eta_{21} &= \beta_0 + \beta_1 \\
\eta_{12} &= \beta_0 + \beta_2 \\
\eta_{22} &= \beta_0 + \beta_2 \\
\eta_{13} &= \beta_0 + \beta_3 \\
\eta_{23} &= \beta_0 + \beta_3
\end{aligned}
\qquad \rightarrow \underline{X} = \qquad
\begin{array}{c} \beta_0 \quad \beta_1 \quad \beta_2 \\ \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{array}\right] \end{array}
$$

In diesem Fall sind die Parameter β_0, β_1, β_2 besonders leicht interpretierbar. Man beachte aber, daß - wenn eine Zelle fehlt - auf die allgemeine Form der zentrierten Effekte zurückgegriffen werden muß, da β_0 sonst nicht als allgemeiner Mittelwert interpretiert werden kann.

Diese Form der Reparametrisierung wird in vielen Programmen zur Varianzanalyse und in ECTA sowie NONMET bei der Analyse von Kontingenztabellen verwendet.

4.2.1.2. Auf eine Kategorie bezogene Effekte (cornered effects)

Eine zweite Möglichkeit der Reparametrisierung besteht einfach darin, den Wert des Parameters einer ausgewählten Kategorie 0 zu setzen.

In der Regel wird die erste Kategorie, z. B. in GLIM oder die letzte, z. B. in ALMO von Holm (1979) verwendet, daher der englische Ausdruck "cornered effects".

Die Matrix $\underline{X}$ nimmt dann in unserem Beispiel zur Kontingenztabelle folgende Form an:

$$\eta_i = \beta_0 + \beta_j \qquad j = 1,2,3 \qquad \beta_1 = 0$$

$$\eta_1 = \beta_0 + \beta_1$$
$$\eta_2 = \beta_0 + \beta_1$$
$$\eta_3 = \beta_0 + \beta_2 \quad \rightarrow \underline{X} =$$
$$\eta_4 = \beta_0 + \beta_3$$
$$\eta_5 = \beta_0 + \beta_3$$

$$\begin{matrix} \beta_0 & \beta_2 & \beta_3 \end{matrix}$$
$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$\rightarrow$ Rg $\underline{X}$ = 3 also von vollem Spaltenrang.

β_0 kann nun nicht mehr als Mittelwert interpretiert werden, sondern ist gleich dem Wert der ersten Kategorie. β_2 und β_3 sind als Abweichungen der zweiten bzw. der dritten Kategorie von der ersten zu interpretieren.

Wir wollen nun den Zusammenhang zwischen cornered effects

$$\beta_0^c,\ \beta_1^c \ \ldots\ \beta_{q_1+1}^c \quad \text{mit} \quad \beta_1^c = 0$$

und zentrierten Effekten

$$\beta_0^z,\ \beta_1^z \ \ldots\ \beta_{q_1+1}^z \quad \text{mit} \quad \sum_{j=1}^{q_1+1} n_j \beta_j = 0$$

untersuchen.

i) einfache zentrierte Effekte: $\sum_{j=1}^{q+1} \beta_j^z = 0$

Die Transformation

$$(4.6)\quad \beta_o^z = \beta_o^c + v,\ v = \frac{1}{q+1} \sum_{j=2}^{q+1} \beta_j^c$$

$$(4.7)\quad \beta_j^z = \beta_j^c - v,\quad j = 1,\dots q+1,\ \beta_1^c = 0$$

liefert das gewünschte Resultat, da

$$\sum_{j=1}^{q+1} \beta_j^z = \sum_{j=1}^{q+1} \beta_j^c - \sum_{l=1}^{q+1} \frac{1}{q+1} \sum_{j=2}^{q+1} \beta_j^c = 0$$

ii) allgemeine zentrierte Effekte: $\sum_{j=1}^{q+1} n_j \beta_j^z = 0$

Hier verwenden wir die Transformation,

$$(4.8)\quad \beta_o^z = \beta_o^c + v,\ v = \frac{1}{m} \sum_{j=2}^{q+1} n_j \beta_j^c,\ m = \sum_{j=1}^{q+1} n_j$$

$$(4.9)\quad \beta_j^z = \beta_j^c - v,\quad j = 1,\dots q+1,\quad \beta_1^c = 0$$

da wiederum

$$\sum_{j=1}^{q+1} n_j \beta_j^z = \sum_{j=1}^{q+1} n_j \beta_j^c - \sum_{j=1}^{q+1} v = 0 \text{ ist.}$$

Es lassen sich also durch leichte Rechnung - die in GLIM auch vom Benutzer durch ein eigenes Unterprogramm - ein sogenanntes Macro - eingebaut werden kann - jederzeit β_j^z aus den β_j^c berechnen.

Für den Unterschied

$\beta_j - \beta_k$, $j,k \in \{1,\ldots q+1\}$ bei beliebiger Reparametrisierung

gilt weiter

$$\beta_j^c - \beta_k^c = (\beta_j^c - v) - (\beta_k^c - v) = \beta_j^z - \beta_k^z$$

d.h. die Unterschiede zwischen zwei Kategorien liegen unabhängig von der gewählten Reparametrisierung eindeutig fest.

Der Vorteil der Reparametrisierung durch cornered effects gegenüber den gewohnten zentrierten Effekten liegt einerseits in der Unabhängigkeit von der Anzahl der Beobachtungen in einer Kategorie, andererseits in der wesentlich einfacheren Behandlung von fehlenden Beobachtungen, die bei zentrierten Effekten die gewohnte Interpretation außerordentlich erschweren. Dies zeigt sich besonders deutlich bei Kreuzklassifikationen und hierarchischen Designs. Ein weiterer Vorteil ist - wie später gezeigt wird - der einfache Übergang von loglinearen zu logistischen Modellen für polytome abhängige Variable.

4.2.2. Kreuzklassifikation

Liegen für ein kombiniertes Merkmal

AxB, AxBxC etc.

Beobachtungen vor mit den Kategorien

A_1B_1, A_2B_1 ... $A_{q_1+1}B_{q_2+1}$, $A_1B_1C_1$ ...,

sprechen wir von einer Kreuzklassifikation. Beispiele sind die üblichen mehrfaktoriellen Versuchspläne, aber auch jede mehrdimensionale Kontingenztabelle.

Wir gehen auf unser Beispiel in 4.2.1.1 zurück

	B_1	B_2	B_3
A_1	y_{11}	y_{12}	y_{13}
A_2	y_{21}	y_{22}	y_{23}

$$\rightarrow \underline{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix}$$

Das saturierte Modell lautet dann

$$\eta_{jl} = \beta_0 + \beta_j^A + \beta_l^B + \beta_{jl}^{AB} \qquad j = 1,2;\ l = 1,2,3$$

β_j^A, β_l^B werden - wie in der klassischen Varianzanalyse - als Haupteffekte, β_{jl}^{AB} als Interaktionseffekte erster Ordnung bezeichnet.

Für 6 Beobachtungen erhalten wir 12 Parameter. Um zu einer eindeutigen Lösung zu gelangen, führen wir 6 lineare Restriktionen ein und konstruieren die entsprechende Designmatrix $\underline{X}$.

i) Zentrierte Effekte

$$\sum_{j=1}^{2} \beta_j^A = 0, \quad \sum_{l=1}^{3} \beta_l^B = 0, \quad \sum_{j=1}^{2} \beta_{jl}^{AB} = 0,\ l = 1,2,3; \quad \sum_{l=1}^{3} \beta_{jl}^{AB} = 0,\ j = 1,2$$

$$\underline{X} = \begin{array}{c} \begin{array}{cccccc} \beta_0 & \beta_1^A & \beta_1^B & \beta_2^B & \beta_{11}^{AB} & \beta_{12}^{AB} \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \end{array}$$

Rg $\underline{X}$ = 6
also von vollem Spaltenrang

ii) Cornered effects

$$\beta_1^A = 0,\ \beta_1^B = 0,\ \beta_{11}^{AB} = \beta_{12}^{AB} = \beta_{13}^{AB} = \beta_{21}^{AB} = 0$$

$$\underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_2^A & \beta_2^B & \beta_3^B & \beta_{22}^{AB} & \beta_{23}^{AB} \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \qquad \text{Rg } \underline{X} = 6$$

In beiden Fällen lassen sich die den Interaktionsparametern zugehörigen Spalten von $\underline{X}$ als Produkte der Spalten der Haupteffekte berechnen, die den jeweiligen Interaktionseffekt definieren.

Das Produkt ist kein inneres Vektorprodukt, sondern ist als elementweises Produkt definiert.

$$\underline{x}_{\beta_{j1}}^{AB} = \underline{x}_{\beta_j}^{A} \cdot \underline{x}_{\beta_1}^{B} \quad \text{mit} \quad \underline{y} \cdot \underline{z} = \begin{bmatrix} y_1 \cdot z_1 \\ y_2 \cdot z_2 \\ \vdots \\ y_n \cdot z_n \end{bmatrix}$$

Es genügt also, die Spalten für die Haupteffekte zu definieren. Daraus ergeben sich die entsprechenden Spalten der Interaktionseffekte durch Vektormultiplikation mit $\cdot$.

Diese Art der Erzeugung von Interaktionen läßt sich auf beliebig viele Kreuzklassifikationen ausdehnen.

Eine elegante mathematische Formulierung mit Hilfe von Kroneckerprodukten wird von Bock (1975, S. 273 ff.) dargestellt.

In der Programmsprache GLIM wird folgende symbolische Darstellung gewählt. (Parametrisierung: cornered effects)

Seien A, B, C ... nominale Variable mit den Ausprägungen

$A_1 \cdots A_{q_1+1}$, $B_1 \cdots B_{q_2+1}$, $C_1 \cdots C_{q_3+1}$, ...

so bedeuten die Zeichen:

+ : A+B zu den Spalten von $\underline{X}$, die den Parametern $\beta_2^A \cdots \beta_{q_1+1}^A$ entsprechen, werden die Spalten, die den Parametern $\beta_2^B + \ldots \beta_{q_2+1}^B$ entsprechen, hinzugefügt.

· : A·B Zu den Spalten von A+B werden die Spalten, die aus der · Multiplikation der A und B zugehörigen Spalten entstehen, hinzugefügt.

* : A*B $\Longleftrightarrow$ A+B+A·B

- : A-B = A+B - A.B . Von $\underline{X}$ werden die A.B entsprechenden Spalten entfernt.

Durch Klammerausdrücke lassen sich die eingeführten Operationen miteinander kombinieren. Die einzelnen Regeln und Rangfolge der Zeichen sind im GLIM Manual (1978) enthalten.

Beispiel:

$$(A+B)*C = A*C+B*C$$
$$= A+C+A.C+B+C+B.C$$
$$= A+B+C+A.C+B.C$$

$$(A+B)*(A+C) = A*A+B*A+A*C+B*C$$
$$= A+B+A.B+C+A.C+B.C$$

$$A*B*C = (A+B+A.B)*C =$$
$$= A+B+C+A.B+A.C+B.C+A.B.C$$

Diese symbolische Schreibweise vereinfacht das Anpassen verschiedener Modelle, da sich im Programm damit interaktiv sofort die Beiträge einzelner Variablen und ihrer Interaktionen berechnen lassen.

4.2.3. Fehlende Beobachtungen

Sowohl bei der klassischen Varianzanalyse als auch in Kontingenztabellen treten häufig Kombinationen auf, für die Beobachtungen fehlen.

Wir bezeichnen Kombinationen, die aus inhaltlichen oder logischen Gründen nicht auftreten können, als fehlende Zellen oder strukturelle Nullen.

Ist die Kombination inhaltlich sinnvoll, aber die Stichprobe zu klein, wird die fehlende Beobachtung als sampling zero bezeichnet. In beiden Fällen können die fehlenden Beobachtungen dazu führen, daß trotz Reparametrisierung der Rang $\underset{n\times p}{X} = k < p$ ist.

Um wieder zu einer eindeutig bestimmten Schätzung zu gelangen, müssen zusätzliche, nämlich p-k lineare Restriktionen eingeführt werden.

Inhaltlich bedeutet das, daß nicht alle gewünschten p Parameter $\beta_1 \ldots \beta_p$ geschätzt werden können, da zu wenig Informationen vorliegen.

Beispiel: Wir gehen wiederum von der Kontingenztabelle

	B_1	B_2	B_3
A_1	y_{11}	y_{12}	y_{13}
A_2	y_{21}	y_{22}	y_{23}

aus, die bei Reparametrisierung durch cornered effects im saturierten Modell wie folgt beschrieben wird

$$Ey_{jl} = \mu_{jl}, \quad \eta_{jl} = \ln\mu_{jl} = \beta_0 + \beta_j^A + \beta_l^B + \beta_{jl}^{AB}$$

mit $\beta_1^A = \beta_1^B = \beta_{11}^{AB} = \beta_{12}^{AB} = \beta_{13}^{AB} = \beta_{21}^{AB} = 0$

$$
\begin{bmatrix} \eta_{11} \\ \eta_{12} \\ \eta_{13} \\ \eta_{21} \\ \eta_{22} \\ \eta_{23} \end{bmatrix} =
\begin{array}{c}
\begin{matrix} \beta_0 & \beta_2^A & \beta_2^B & \beta_3^B & \beta_{22}^{AB} & \beta_{23}^{AB} \end{matrix} \\
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}
\end{array}
\cdot
\begin{bmatrix} \beta_0 \\ \beta_2^A \\ \beta_2^B \\ \beta_3^B \\ \beta_{22}^{AB} \\ \beta_{23}^{AB} \end{bmatrix}
$$

Wenn nun eine Beobachtung fehlt, z. B. y_{12}, wird die entsprechende Zeile von $\underline{\eta}$ und $\underline{X}$ gestrichen, die Matrix $\underline{X}$ ohne die zweite Zeile ist dann

$$
\underset{5x6}{X^*} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}
$$

Die 5. Spalte - die mit β_{22}^{AB} multipliziert wird - ist dann identisch mit der 3. Spalte - die mit β_2^B multipliziert wird. Der Rang der Matrix ist daher Rang $\underline{X}^* = 5$. Als Folge der fehlenden Beobachtung können wir β_{22}^{AB} nicht mehr schätzen, wir setzen diesen Wert daher = 0. Im Programmsystem GLIM wird ein Wert, der auf Grund fehlender Beobachtungen nicht mehr berechnet werden kann, als aliased ausgewiesen.

Liegt eine Matrix $\underline{X}$ vor oder wird sie mit Hilfe der oben genannten Produkte aus den Spalten für die Haupteffekte aufgebaut, muß für jede Spalte - etwa durch Berechnung von $|\underline{X}'\underline{X}|$ - geprüft werden, ob sie von den vorhergehenden Spalten auf Grund fehlender Beobachtungen linear abhängig ist, oder nicht. Liegt lineare Abhängigkeit vor, wird diese Spalte weggelassen und der entsprechende Parameter = 0 gesetzt. Diese Prozedur wird sowohl in GLIM als auch in ALMO von Holm (1979) durchgeführt.

4.2.4. Hierarchische Varianzanalyse

Häufig tritt der Fall auf, daß innerhalb einer Kategorie von A mehrere Unterkategorien

$$B_1^{A_1} \ldots B_{r_1}^{A_1}, \quad B_1^{A_2} \ldots B_{r_2}^{A_2},$$

usw. auftreten, die je nach Kategorie von A verschieden sein und nur nach Eintritt einer Kategorie von A auftreten können. Sie entsprechen daher nicht einer Kreuzklassifikation. In diesem Fall spricht man von hierarchischer Varianzanalyse. Als Beispiel kann man sich die gewählten Berufe nach Abschluß eines bestimmten Schultyps vorstellen. Die Effekte von B werden, da sie nur nach Eintritt eines A_j auftreten, als durch A_j bedingte oder konditionale Effekte bezeichnet.

$$(4.10)\quad \eta_{jl} = \beta_0 + \beta_j^A + \beta_{jl}^{AB} \qquad \begin{matrix} j = 1,2 \ldots q \\ l = 1,2 \ldots r \end{matrix}$$

$A_1 \ldots A_q$ sind die Kategorien von A

r ist die maximale Zahl von Unterkategorien

$B_j^{A_j} \ldots B_{r_j}^{A_j}$, die in einer Kategorie von A auftreten können.

Man beachte, daß die Kategorien $B_1^{A_1}$ und $B_1^{A_j}$ nicht gleich sein müssen.

Da die Kategorien von B nur nach Eintritt von A auftreten können, fallen die Haupteffekte von B weg.

Man muß allerdings bei Gleichung (4.10) beachten, daß die Reparametrisierung erst nach dem Ausrechnen der Produkte für die Interaktionseffekte vorgenommen werden kann, da sonst z. B. bei cornered effects die Effekte der Subgruppen innerhalb von A_1 alle 0 gesetzt werden.

Beispiel: FUr die Kategorien A_1, A_2, A_3 einer Variablen A finden wir die Unterkategorien

$$B_1^{A_1},\ B_2^{A_1},\ B_1^{A_2},\ B_2^{A_2},\ B_3^{A_2},\ B_1^{A_3},\ B_2^{A_3}$$

denen die Interaktionsparameter

$$\beta_{11}^{AB},\ \beta_{12}^{AB},\ \beta_{21}^{AB},\ \beta_{22}^{AB},\ \beta_{23}^{AB},\ \beta_{31}^{AB},\ \beta_{32}^{AB}$$

entsprechen.

Die maximale Anzahl der Subkategorien ist r = 3. Wenn wir daher vor der Reparametrisierung gemäß

$$\eta_{jl} = \beta_0 + \beta_j^A + \beta_{jl}^{AB} \qquad j = 1,2,3;\quad r = 1,2,3$$

die Matrix $\underline{X}$ aufbauen, erhalten wir, wenn wir für jede Kombination von A und B eine Beobachtung z. B. in einer Kontingenztabelle annehmen:

$$\underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_1^A & \beta_2^A & \beta_3^A & \beta_{11}^{AB} & \beta_{12}^{AB} & \beta_{13}^{AB} & \beta_{21}^{AB} & \beta_{22}^{AB} & \beta_{23}^{AB} & \beta_{31}^{AB} & \beta_{32}^{AB} & \beta_{33}^{AB} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Die Parameter β_{13}^{AB}, β_{33}^{AB} fallen auf Grund fehlender Subkategorien sofort weg. Ihnen entspricht jeweils ein Nullvektor als Spalte, so daß sie wie fehlende Beobachtungen behandelt werden.

Um die Matrix auf vollen Spaltenrang zu bringen, führen wir folgende Reparametrisierung mit cornered effects durch

$$\beta_1^A = \beta_{11}^{AB} = \beta_{21}^{AB} = \beta_{31}^{AB} = 0$$

Die Reparametrisierung entspricht der üblichen Reparametrisierung von Interaktionseffekten, nur läßt sich durch den Wegfall der bei Kreuzklassifikation üblichen Schätzung der Effekte von B auch der Parameter β_{12}^{AB} schätzen.

Wir erhalten damit folgende Matrix $\underline{X}$ nach Reparametrisierung

$$\underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_2^A & \beta_3^A & \beta_{12}^{AB} & \beta_{22}^{AB} & \beta_{23}^{AB} & \beta_{32}^{AB} \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

GLIM sieht für die hierarchische Varianzanalyse einen Operator (nesting operator) vor, der den oben dargestellten Aufbau der Matrix $\underline{X}$ bewirkt.

Sei A eine Variable mit q Kategorien und r die maximale Anzahl der Subkategorien in A_j, $j = 1,\ldots q$, so gilt die Formel

/: A/B = A+A.B

A/B/C = A+A.B+A.B.C

Man beachte, daß diese hierarchische Analyse mit der oben angegebenen Kreuzklassifikation kombiniert werden kann.

Schließlich können auch die Interaktionseffekte der Kreuzklassifikation als einfache lineare Funktion der Parameter bei hierarchischer Varianzanalyse begriffen werden. Wir zeigen dies am Beispiel von zwei Variablen A,B mit Kategorien A_1, A_2; B_1, B_2 die zunächst als hierarchisch geordnet, dann als Kreuzklassifikation aufgefaßt werden.

Beispiel:

Reparametrisierung durch zentrierte Effekte

Konditionale Effekte

$$\underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_1^A & \beta_{11}^{AB} & \beta_{21}^{AB} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix} \end{array}$$

Kreuzklassifikation

$$\underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_1^A & \beta_1^B & \beta_{11}^{AB} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \end{array}$$

Die dem Interaktionseffekt β_{11}^{AB} der Kreuzklassifikation zugeordnete Spalte läßt sich als Differenz der den konditionalen Effekten $\beta_{11}^{AB} - \beta_{21}^{AB}$ zugeordneten Spalten schreiben.

Reparametrisierung durch cornered effects

Konditionale Effekte

$$\underline{X} = \begin{matrix} \beta_0 & \beta_2^A & \beta_{12}^{AB} & \beta_{22}^{AB} \\ \end{matrix}$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

Kreuzklassifikation

$$\underline{X} = \begin{matrix} \beta_0 & \beta_2^A & \beta_2^B & \beta_{22}^{AB} \\ \end{matrix}$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Die Spalte Interaktionseffekts β_{22}^{AB} der Kreuzklassifikation ist gleich der Spalte des konditionalen Effekts β_{22}^{AB}.

4.2.5. Kovarianzanalyse

Varianzanalyse und Regressionsanalyse lassen sich nun verbinden, indem wir in $\underline{X}$ sowohl Dummy-Variable als auch quantitative Variable zulassen.

Ein einfaches Beispiel - wie bereits in 1.2 dargestellt - ist die Abhängigkeit der Berufstätigkeit verheirateter Frauen von der Variablen, ob sie Kinder haben oder nicht (A_1, A_2) und vom Einkommen des Mannes, gemessen z. B. in 5 Stufen z_1, z_2, z_3, z_4, z_5.

Sei y der Prozentsatz an nicht erwerbstätigen Frauen für die einzelnen Stufen von z, dann verwenden wir das Modell:

y_{jl} ist der Prozentsatz in Kategorie A_j und Einkommensstufe z_l.

$$Ey_{jl} = \mu_{jl} \qquad \eta_{jl} = \ln \frac{\mu_{jl}}{1-\mu_{jl}} \qquad j = 1,2 \qquad l = 1,\ldots 5$$

Reparametrisierung $\beta_1^A = 0$ ergibt das Modell $\underline{n} = \underline{X\beta}$ mit

$$\underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_2^A & \beta^Z \end{matrix} \\ \begin{bmatrix} 1 & 0 & z_1 \\ 1 & 0 & z_2 \\ 1 & 0 & z_3 \\ 1 & 0 & z_4 \\ 1 & 0 & z_5 \\ 1 & 1 & z_1 \\ 1 & 1 & z_2 \\ 1 & 1 & z_3 \\ 1 & 1 & z_4 \\ 1 & 1 & z_5 \end{bmatrix} \end{array}$$

Wir haben dabei angenommen, daß der Regressionseffekt β_3 für beide Gruppen von A gleich ist. Inhaltlich bedeutet das, daß der Effekt des Einkommens des Mannes auf die Wahrscheinlichkeit, nicht erwerbstätig zu sein, unabhängig davon ist, ob eine Frau Kinder hat oder nicht.

Wenn diese Hypothese nicht gerechtfertigt erscheint, können wir einen Interaktionsparameter hinzufügen, indem wir die Spalte von β_2^A mit der Spalte von β^Z, also mit den z Werten multiplizieren.

Wir erhalten dann als Matrix $\underline{X}$ für das obige Beispiel

$$\underline{X} = \begin{array}{c} \begin{matrix} \beta_0 & \beta_2^A & \beta^Z & \beta_2^{A.Z} \end{matrix} \\ \begin{bmatrix} 1 & 0 & z_1 & 0 \\ 1 & 0 & z_2 & 0 \\ 1 & 0 & z_3 & 0 \\ 1 & 0 & z_4 & 0 \\ 1 & 0 & z_5 & 0 \\ 1 & 1 & z_1 & z_1 \\ 1 & 1 & z_2 & z_2 \\ 1 & 1 & z_3 & z_3 \\ 1 & 1 & z_4 & z_4 \\ 1 & 1 & z_5 & z_5 \end{bmatrix} \end{array}$$

Für die erste Gruppe lautet nun die Gleichung

$$\eta_{1l} = \beta_0 + \beta z_l \qquad l = 1, \ldots 5$$

Für die zweite Gruppe lautet die Gleichung

$$\eta_{2l} = \beta_0 + \beta_2^A + (\beta^Z + \beta_2^{A.Z}) z_l \qquad l = 1, \ldots 5$$

β_2^A gibt an, um wieviel sich die Wahrscheinlichkeit im Vergleich zur ersten Gruppe erhöht hat, $\beta_2^{A.Z}$ gibt an, um wieviel sich der Regressionseffekt von Z im Vergleich zur zweiten Gruppe erhöht hat.

Mit Hilfe des Produktoperators . lassen sich daher in GLIM nicht nur Interaktionen von qualitativen Variablen, sondern auch von quantitativen mit sich oder mit qualitativen Variablen bilden.

A.X — alle Spalten, die den Ausprägungen von A entsprechen, werden mit der Spalte der quantitativen Variablen X multipliziert.

A*B*X = A+B+X+A.B+A.X+B.X+A.B.X

Es wird angenommen, daß sich die Regressionskoeffizienten für jede Kategorie von A und B und für jede Kombination A.B unterscheiden.

X.X = X^2. Es werden die Werte x_i^2 i = 1,... n gebildet.

X+X.X — Sowohl lineare als auch quadratische Regressionseffekte werden gebildet.

4.3. Freiheitsgrade

Die Anzahl der Freiheitsgrade für den Test der Anpassung des Modells an die Daten ist gegeben durch

(4.11) df = n-p

n = Anzahl der Beobachtungen

p = Rang der Matrix $\underline{X}$

4.4. Orthogonalisierung und Standardisierung

Bekanntlich ändern sich sowohl bei der Regression als auch bei der Varianzanalyse die geschätzten Parameter $\underline{b}$, wenn man Variable wegläßt oder hinzufügt.

Ausnahmen bilden nur die sogenannten orthogonalen Designs der Varianzanalyse, in denen für jede Kombination von zwei oder mehr Variablen gleiche Zellenbesetzung angenommen wird. Andererseits ist es häufig wünschenswert, Variable so zu definieren, daß sich ihr Regressionskoeffizient nicht mehr ändert, wenn weitere Variable zugelassen werden.

Dies ist der Fall, wenn bei einer Variablen $\underline{x}_j$ alle Einflüsse der vorhergehenden Variablen $\underline{x}_1 \ldots \underline{x}_{j-1}$ eliminiert sind. Die transformierte Variable wird wie folgt berechnet:

$$\underline{x}_j^* = \underline{x}_j - \beta_0 - \beta_1\underline{x}_1 - \beta_2\underline{x}_2 \ldots - \beta_{j-1}\underline{x}_{j-1}$$

wobei $\beta_1 \ldots \beta_{j-1}$ Regressionskoeffizienten einer kleinsten Quadrate-Lösung mit x_j als abhängiger Variablen sind. $\underline{x}_j^*$ läßt sich daher als Fehlervektor von $\underline{x}_j$ bezüglich der anderen Variablen interpretieren. Der Regressionskoeffizient von $\underline{x}_j^*$ gibt daher an, wieviel Erklärungskraft $\underline{x}_j$ aufweist, wenn der Einfluß aller vorhergehenden Variablen auspartiellert wurde. Da es sich im allgemeinen Fall um gewichtete Regression handelt, muß bei obiger Formel die sich bei jeder Iteration ändernde Gewichtung berücksichtigt werden.

Formal bedeutet die Orthogonalisierung, daß die geschätzten Koeffizienten $\underline{b}$ untereinander mit 0 korreliert sind. Da, wie wir in 3.2. gezeigt haben, asymptotisch gilt

$$(3.8)\quad \underline{b} \sim N(\underline{\beta}, (\underline{X}'\underline{W}\underline{X})^{-1} = N(\underline{\beta}, \hat{\underline{\Sigma}}_b)$$

müssen die Variablen x_j, $j = 1,\ldots p$ so transformiert werden, daß gilt:

$$\hat{\underline{\Sigma}}_u = \underline{I}$$

wenn $\underline{u}$ die geschätzten Regressionskoeffizienten der transformierten Variablen x_j, $j = 1,\ldots p$ sind.

Da sich $\underline{W}$ - außer im Fall der link Funktion $\eta = \mu$ - bei jeder Iteration ändert, ist es unökonomisch, bei jeder Iteration die Variablen x_j $j = 1,\ldots p$ etwa durch Regressionsrechnung zu orthogonalisieren.

Wir berechnen direkt die Schätzwerte u_j, $j = 1,\ldots p$ der orthogonalen Koeffizienten γ_j, $j = 1,\ldots p$, indem wir folgende Überlegungen anwenden.

$\hat{\underline{\Sigma}}_b$ ist als Kovarianzmatrix positiv definit. Sie läßt sich daher als Produkt einer Dreiecksmatrix mit ihrer Transponierten, der sogenannten Choleskymatrix - die eindeutig bestimmt ist - darstellen (vgl. Bock(1975, S. 85 ff.)).

(4.12) $\hat{\underline{\Sigma}}_b = \underline{S}\underline{S}'$

Wir multiplizieren nun $\underline{b}$ mit $\underline{S}^{-1}$ und erhalten

(4.13) $\underline{S}^{-1}\underline{b} = \underline{u}$ mit

(4.14) $E\underline{u} = E\underline{S}^{-1}\underline{b} = \underline{S}^{-1}\underline{\beta} = \underline{\gamma}$ (orthogonale Koeffizienten)

(4.15) $\hat{\underline{\Sigma}}_u = \underline{S}^{-1}\hat{\underline{\Sigma}}_b\underline{S}^{-1'} = \underline{S}^{-1}\underline{S}\underline{S}'\underline{S}^{-1'} = \underline{I}$

Somit gilt

(4.16) $\underline{u} \sim N(\underline{\gamma},\underline{I})$

Damit gilt unter der H_o: $\gamma_j = 0$, daß $u_j^2 \ldots \chi^2$ verteilt ist mit einem Freiheitsgrad und unter H_o: $\gamma_{j+1} = \gamma_{j+2} \cdots \gamma_k = 0$, daß $\sum_{l=j+1}^{k} u_l^2 \ldots \chi^2$ verteilt ist mit k-j Freiheitsgraden.

Da der Einfluß aller vorhergehenden Variablen eliminiert ist, gestatten es Tests der oben angegebenen H_o Hypothesen, sofort zu Überprüfen, ob einzelne Variable bzw. Kategorien noch einen signifikanten Beitrag zur Erklärung leisten, ohne daß ein neues Modell gerechnet werden muß. Allerdings muß auf Grund der Orthogonalisierung die Reihenfolge der Variablen inhaltlich sinnvoll festgelegt werden; d.h. die zu Überprüfende Variable ist als letzte Variable anzuordnen.

Diese Orthogonalisierung läßt sich in GLIM leicht durch das Schreiben eines Unterprogramms durchführen.

Neben der Orthogonalisierung ist für den Sozialforscher häufig auch die Standardisierung von Interesse, um die Einflußstärke verschiedener Variablen vergleichen zu können. In der Regressionsrechnung erfolgt eine Standardisierung auf das Intervall (-1,1), indem sowohl die abhängige als auch die unabhängige Variablen auf Mittelwert 0 und Standardabweichung 1 transformiert werden. Sind die Regressionskoeffizienten bekannt, gilt

$$(4.17)\quad \beta_j^* = b_j \frac{s_{x_j}}{s_y}$$

β_j^* ist standardisierter Regressionskoeffizient

b_j ist nicht stanardisierter Regressionskoeffizient

s_{x_j} ist geschätzte Standardabweichung von x_j

s_y ist geschätzte Stanardabweichung von y

Da s_y für alle $j = 1,\ldots p$ gleich ist, genügt es zum Vergleich der Größe der Effekte nur mit s_{x_j} zu standardisieren.

Da im allgemeinen Modell eine gewichtete Regression durchgeführt wird, müssen wir die Varianz der gewichteten Variablen x_j^* mit

$$(4.18)\quad x_{ij}^* = x_{ij} w_i^{-\frac{1}{2}} \qquad i = 1,\ldots n$$

berechnen. Da in GLIM die Gewichte w_i bekannt sind, bietet die Berechnung der bezüglich der Varianzen von x_j^* standardisierten Regressionskoeffizienten mit Hilfe eines Macros keine Schwierigkeit.

$$(4.19)\quad \beta_j^* = b_j s_{x_j^*} \quad \text{mit} \quad s_{x_j^*} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij}^* - \bar{x}_j^*).$$

5. Polytome abhängige Variable

Ein spezielles Problem stellen häufig Modelle für abhängige nominale Variable mit mehr als zwei Kategorien dar.

Es gibt hier mehrere Möglichkeiten, dieses Analogon zur multivariaten Regression zu betrachten. Sie werden in der Regel unter der Rubrik "multivariate logits" abgehandelt, da sie Verallgemeinerungen der in 2.3 eingeführten logits darstellen.

Verschiedene Möglichkeiten, multivariate logits zu definieren, wollen wir an folgendem Beispiel demonstrieren:

Die abhängige Variable C mit Ausprägung C_k
$k = 1,\ldots K$ wird von zwei unabhängigen Variablen A mit
A_i, $i = 1,\ldots I$ und B, B_j $j = 1,\ldots J$ beeinflußt. Das loglineare Model A+B+C+A.C+B.C ergibt eine gute Anpassung an die Daten.

In Parametern lautet das Modell

$$\eta_{ijk} = \ln\mu_{ijk} = \beta_o + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{ik}^{AC} + \beta_{jk}^{BC}$$

Wenn wir C als abhängige Variable betrachten, können wir z. B. jede Kategorie C_k $k = 2,\ldots K$ mit der ersten Kategorie C_1 vergleichen.

$$(5.1)\quad \ln \frac{\mu_{ijk}}{\mu_{ij1}} = \eta_{ijk} - \eta_{ij1} = \beta_0 + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{ik}^{AC} + \beta_{jk}^{BC}$$

$$- \beta_0 - \beta_i^A - \beta_j^B - \beta_1^C - \beta_{ij}^{AB} - \beta_{i1}^{AC} - \beta_{j1}^{BC}$$

$$= \underbrace{\beta_k^C - \beta_1^C}_{\lambda} + \underbrace{\beta_{ik}^{AC} - \beta_{i1}^{AC}}_{\lambda_i^A} + \underbrace{\beta_{jk}^{AC} - \beta_{j1}^{BC}}_{\lambda_j^B}$$

$$= \lambda + \lambda_i^A + \lambda_j^B$$

Diese Darstellung ist besonders günstig, wenn wir für das loglineare Modell cornered effects als Reparametrisierung gewählt haben. Da in diesem Fall $\beta_1^C = \beta_{i1}^{AC} = \beta_{j1}^{BC} = 0$ sind, gilt

$$(5.2)\quad \eta_{ijk} - \eta_{ij1} = \beta_k^C + \beta_{ik}^{AC} + \beta_{jk}^{BC} \qquad k = 2,\ldots K$$

Wurde als Reparametrisierung das Nullsetzen der Effekte der jeweils letzten Ausprägung gewählt - wie etwa im Programm ALMO von Holm - empfiehlt es sich, die K-1 logits

$$(5.3)\quad \ln \frac{\mu_{ijk}}{\mu_{ijK}} = \eta_{ijk} - \eta_{ijK} \qquad k = 1,2\ldots K-1$$

zu betrachten.

Wir erläutern die Interpretation an den in 1.2. eingeführten Daten. Die Variablen A (Ausbildung der Frau), B (Kinder) und C (Art der Beschäftigung) sind gleich wie in 1.2 definiert. Im Unterschied zu 1.2. betrachten wir jetzt

C mit allen drei Ausprägungen

C1 Nicht erwerbstätig

C2 Un/angelernte Arbeiterin

C3 Ausführende Angestellte und Beamte

als abhängige Variable. Um die Tabelle zu vereinfachen, lassen wir das Einkommen des Mannes als unabhängige Variable unberücksichtigt.

Die Tabelle der einzelnen Kombinationen für A,B,C hat folgende Gestalt

Tabelle 5.1: Art der Erwerbstätigkeit von Frauen
N = Häufigkeit in der Kombination ABC

A	B	C	N
1	1	1	593
1	1	2	731
1	1	3	175
1	2	1	4697
1	2	2	2045
1	2	3	325
1	3	1	4657
1	3	2	713
1	3	3	118
2	1	1	478
2	1	2	190
2	1	3	1194
2	2	1	4450
2	2	2	609
2	2	3	1324
2	3	1	5692
2	3	2	247
2	3	3	708
3	1	1	28
3	1	3	34
3	2	1	142
3	2	2	1
3	2	3	24
3	3	1	317
3	3	3	25

Man beachte, daß die Kombinationen 3 1 2 und 3 3 2 fehlen, hier liegen keine Beobachtungen vor, wir behandeln sie als strukturelle Nullen.

Da uns nur der Einfluß von A und B auf C interessiert, nicht aber der Zusammenhang von A und B, geben wir die Randverteilung von AB als gegeben vor und passen zunächst das Modell

$$Ey_{ijk} = \mu_{ijk},\ \eta_{ijk} = \ln\mu_{ijk}$$

$$(5.4)\quad \eta_{ijk} = \beta_0 + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{ik}^{AC} + \beta_{jk}^{BC},\ i,j,k = 1,2,3$$

an.

Auf Grund der großen Stichprobe erhalten wir

$$D(c,f) = 36.53 \quad \text{mit} \quad df = 6,$$

so daß dieses Modell mit $\alpha = 0.05$ verworfen werden muß.

Wir berechnen daher die Parameter des saturierten Modells

$$(5.5)\quad \eta_{ijk} = \beta_0 + \beta_i^A + \beta_j^B + \beta_k^C + \beta_{ij}^{AB} + \beta_{ik}^{AC} + \beta_{jk}^{BC} + \beta_{ijk}^{ABC} \quad i,j,k=1,2,3$$

und erhalten bei Reparametrisierung durch cornered effects bezogen auf die jeweils erste Kategorie:

Tabelle 5.2.: Geschätzte Parameter des saturierten Modells 5.5

```
ERROR POISSON  LINK LOG

LINEAR PREDICTOR
<GM A B C A.B A.C B.C A.B.C
```

	ESTIMATE	S.E.	PARAMETER
1	6.385	.4107E-01	<GM
2	-.2156	.6147E-01	A(2)
3	-3.053	.1934	A(3)
4	2.069	.4358E-01	B(2)
5	2.061	.4360E-01	B(3)
6	.2092	.5527E-01	C(2)
7	-1.220	.8603E-01	C(3)
8	.1616	.6493E-01	A(2).B(2)
9	.4163	.6457E-01	A(2).B(3)
10	-.4459	.2113	A(3).B(2)
11	.3658	.2019	A(3).B(3)
12	-1.132	.1020	A(2).C(2)
13	2.136	.1016	A(2).C(3)
14	-4.124	1.003	A(3).C(2)
15	1.415	.2693	A(3).C(3)
16	-1.041	.6129E-01	B(2).C(2)
17	-1.450	.1034	B(2).C(3)
18	-2.086	.6835E-01	B(3).C(2)
19	-2.455	.1268	B(3).C(3)
20	-.2551E-01	.1139	A(2).B(2).C(2)
21	-.6773	.1208	A(2).B(2).C(3)
22	-.1290	.1275	A(2).B(3).C(2)
23	-.5448	.1436	A(2).B(3).C(3)
24	ZERO	ALIASED	A(3).B(2).C(2)
25	-.5215	.3529	A(3).B(2).C(3)
26	ZERO	ALIASED	A(3).B(3).C(2)
27	-.2791	.3527	A(3).B(3).C(3)

Auf Grund der fehlenden Beobachtungen werden $\beta^{ABC}_{322} = \beta^{ABC}_{332} = 0$ gesetzt und als nicht schätzbar durch ALIASED ausgewiesen.

Zur Analyse der Wirkung von A,B auf C betrachten wir, wie oben vorgeschlagen, zunächst das Verhältnis der Wahrscheinlichkeit, Un/angelernte Arbeiterin, zur Wahrscheinlichkeit, nicht erwerbstätig zu sein

$$\eta_{ij2} - \eta_{ij1} = \beta_2^C + \beta_{i2}^{AC} + \beta_{j2}^{BC} + \beta_{ij2}^{ABC}$$

$\beta_2^C = .2092$

$\beta_{22}^{AC} = -1.132$

$\beta_{32}^{AC} = -4.124$

$\beta_{22}^{BC} = -1.041$

$\beta_{32}^{BC} = -2.086$

Interpretiert man nun die Haupteffekte, wird die Wahrscheinlichkeit - im Vergleich zu Nichterwerbstätigen - an/ungelernte Arbeiterin zu sein, bei mittlerer bzw. höherer Ausbildung und bei Versorgung von Kindern wesentlich vermindert. Vergleicht man die Größe dieser Haupteffekte im logit Modell mit den Interaktionseffekten, erkennt man sofort die relativ geringe Bedeutung der Interaktionswirkung von A.B auf C.

Ähnlich verhält es sich mit den logits, die das Verhältnis der Wahrscheinlichkeit, Angestellte oder Beamte zu sein zur Wahrscheinlichkeit, nicht erwerbstätig zu sein, angeben.

$$\eta_{ij3} - \eta_{ij1} = \beta_3^C + \beta_{i3}^{AC} + \beta_{j3}^{BC} + \beta_{ij3}^{ABC}$$

$\beta_3^C = -1.220$

$\beta_{23}^{AC} = 2.136$

$\beta_{33}^{AC} = 1.415$

$\beta_{23}^{BC} = -1.450$

$\beta_{33}^{BC} = -2.455$

Wie zu erwarten war, wirkt B in gleicher Weise auf C3 wie auf C2. Die Wahrscheinlichkeit der Berufstätigkeit vermindert sich, wenn Kinder zu versorgen sind.

Genau in die andere Richtung wirkt jedoch Variable A auf C3. Die Wahrscheinlichkeit, als Angestellte oder Beamte tätig zu sein, wird bei mittlerer oder höherer Ausbildung beträchtlich erhöht.

Wenn die Kategorien von C geordnet sind, ist es sinnvoll, die sogenannten "continuation ratios" von Fienberg (1977, S. 86 ff.) zu verwenden. Wir gehen dabei von folgender Überlegung aus. Wenn die Kategorien von C geordnet sind, betrachten wir jeweils die Wahrscheinlichkeit, in Kategorie C_k zu kommen im Verhältnis zur Wahrscheinlichkeit, in eine der Kategorien $C_{k+1}, \ldots C_K$ zu fallen. Da $C_1 \ldots C_K$ geordnet sind, interessiert jeweils nur der Vergleich zu den nächst höheren Kategorien. Dies läßt sich als Folge von K-1 logits auffassen.

$$(5.6)\quad \eta_{ij}^{(1)} = \ln \frac{\mu_{ij1}}{\sum_{k=2}^{K} \mu_{ijk}}$$

$$\eta_{ij}^{(2)} = \ln \frac{\mu_{ij2}}{\sum_{k=3}^{K} \mu_{ijk}}$$

$$\vdots$$

$$\eta_{ij}^{(K-1)} = \ln \frac{\mu_{ijK-1}}{\mu_{ijK}}$$

An Stelle eines loglinearen Modells für $\ln\mu_{ijk}$ lassen sich mit obiger Definition (K-1) logit Modelle für $n_{ij}^{(k)}$ $k = 1,\ldots K-1$ rechnen. Dabei ist zu beachten, daß sich für jedes logit Modell die Häufigkeiten in den einzelnen Beobachtungen ändern.

1. Modell: $y_{ij+}^{(1)} = \sum_{k=1}^{K} y_{ijk}$

2. Modell: $y_{ij+}^{(2)} = \sum_{k=2}^{K} y_{ijk}$

.
.
.

K-1. Modell: $y_{ij+}^{(K-1)} = y_{ij(K-1)} + y_{ijK}$

Die Parameter der einzelnen logit Modelle sind bei dieser Art der Berechnung voneinander asymptotisch unabhängig, da - wie im Anhang gezeigt wird - sich für jede Kombination $A_i B_j$ die Likelihoodfunktion der Multinomialverteilung als Produkt der Likelihoodfunktionen binomialverteilter Zufallsvariablen mit den in 5.4 definierten kanonischen Parametern n_{ij} schreiben läßt.

Diese und ähnliche Definitionen multivariater logits bzw. analoger Funktionen lassen sich mit speziellen linearen Modellen zur dem Meßniveau angepaßten Analyse von abhängigen ordinalen Variablen heranziehen.

Im Beispiel in 1.2 haben wir das Verhältnis der Wahrscheinlichkeiten nicht erwerbstätig gegen erwerbstätig als abhängige Variable untersucht. Wir führen nun eine Analyse der Wirkung von Ausbildung, Kinder und Einkommen des Mannes auf Art der Erwerbstätigkeit mit continuation ratios durch, indem wir folgende logits bilden.

μ_{i1}, μ_{i2}, μ_{i3} seien die erwarteten relativen Häufigkeiten der Kategorien C1, C2, C3.

$$(5.7)\quad n_i^{(1)} = \ln \frac{\mu_{i1}}{\mu_{i2}+\mu_{i3}} \qquad i = 1,2\ldots n$$

$$(5.8)\quad n_i^{(2)} = \ln \frac{\mu_{i2}}{\mu_{i3}} \qquad i = 1,2\ldots n$$

Die erste logit Analyse wurde bereits in 1.2 durchgeführt.

Die zweite logit Analyse geht von folgender Tabelle aus.

Tabelle 5.3: Art der Erwerbstätigkeit von Frauen

R2 = Anzahl der an/ungelernten Arbeiterinnen

N2 = Anzahl der erwerbstätigen Frauen in der Kombination AßX

A,B,X,Z folgen den Bezeichnungen der Tabelle 1.1

X	A	B	R2	N2	Z
5	1	1	0015	0016	04.5
5	1	2	0042	0044	04.5
5	1	3	0012	0014	04.5
5	2	1	0006	0011	04.5
5	2	2	0011	0022	04.5
5	2	3	0003	0009	04.5
4	1	1	0214	0251	07.0
4	1	2	0481	0515	07.0
4	1	3	0172	0186	07.0
4	2	1	0044	0170	07.0
4	2	2	0095	0201	07.0
4	2	3	0033	0099	07.0
4	3	1	0000	0002	07.0
4	3	3	0000	0001	07.0
3	1	1	0424	0516	10.0
3	1	2	1285	1473	10.0
3	1	3	0448	0521	10.0
3	2	1	0114	0671	10.0
3	2	2	0391	1052	10.0
3	2	3	0149	0493	10.0
3	3	1	0000	0012	10.0
3	3	2	0000	0005	10.0
3	3	3	0000	0002	10.0
2	1	1	0071	0107	15.0
2	1	2	0230	0319	15.0
2	1	3	0078	0104	15.0
2	2	1	0025	0439	15.0
2	2	2	0105	0560	15.0
2	2	3	0058	0302	15.0
2	3	1	0000	0013	15.0
2	3	2	0001	0012	15.0
2	3	3	0000	0014	15.0
1	1	1	0007	0016	20.0
1	1	2	0007	0019	20.0
1	1	3	0003	0006	20.0
1	2	1	0001	0093	20.0
1	2	2	0007	0098	20.0
1	2	3	0004	0052	20.0
1	3	1	0000	0007	20.0
1	3	2	0000	0008	20.0
1	3	3	0000	0008	20.0

Mit der gleichen Prozedur wie in 1.1. erhalten wir als gut passendes sparsames Modell

$$(5.9)\quad \eta_{ij}^{(2)} = \beta_0 + \beta_i^A + \beta_j^B + \beta_{ij}^{AB} + \beta' z \qquad i,j = 1,2,3 \text{ mit}$$

$$D(c,f) = 26.81 \quad df = 31$$

Als Ergebnis der Schätzung erhalten wir

Tabelle 5.4.: Schätzungen und Standardabweichungen der Parameter in 5.7., sowie standardisierte Fehler.

	ESTIMATE	S.E.	PARAMETER
1	3.249	.1351	<GM
2	-3.114	.1173	A(2)
3	-12.80	37.32	A(3)
4	.4454	.1056	B(2)
5	.3883	.1329	B(3)
6	-.1778	.9829E-02	Z
7	.6344	.1414	A(2).B(2)
8	.4285	.1725	A(2).B(3)
9	8.516	37.34	A(3).B(2)
10	.3078E-01	57.02	A(3).B(3)

UNIT	OBSERVED	OUT OF	FITTED	RESIDUAL
1	15	16	14.73	.2516
2	42	44	41.69	.2077
3	12	14	13.23	-1.432
4	6	11	3.737	1.441
5	11	22	13.25	-.9806
6	3	9	4.842	-1.231
7	214	251	221.2	-1.405
8	481	515	474.1	1.126
9	172	186	170.4	.4160
10	44	170	42.17	.3256
11	95	201	99.03	-.5684
12	33	99	42.32	-1.893
13	0	2	.4104E-04	-.6406E-02
14	0	1	.3120E-04	-.5586E-02
15	424	516	419.6	.4931
16	1285	1473	1284.	.6899E-01
17	448	521	450.8	-.3586
18	114	671	108.8	.5450
19	391	1052	381.8	.5881
20	149	493	150.2	-.1142
21	0	12	.1445E-03	-.1202E-01
22	0	5	.4291	-.6851
23	0	2	.3661E-04	-.6051E-02
24	71	107	68.65	.4729
25	230	319	234.9	-.6285
26	78	104	75.43	.5649
27	25	439	32.35	-1.343
28	105	560	106.3	-.1382
29	58	302	46.09	1.906
30	0	13	.6434E-04	-.8021E-02
31	1	12	.4459	.8456
32	0	14	.1054E-03	-.1026E-01
33	7	16	6.784	.1093
34	7	19	10.16	-1.453
35	3	6	3.123	-.1004
36	1	93	2.946	-1.152
37	7	98	8.609	-.5743
38	4	52	3.585	.2273
39	0	7	.1424E-04	-.3774E-02
40	0	8	.1250	-.3563
41	0	8	.2476E-04	-.4975E-02

Bei der Interpretation der Parameter können wir sofort β_3^A, β_{32}^{AB}, β_{33}^{AB} eliminieren, die Anzahl der Fälle ist offensichtlich in A3 so klein, daß keine Aussagen möglich sind. Dies läßt sich an der Größe der Standardabweichungen erkennen.

Die Bedeutung von Ausbildung und Kinder kehrt sich im Vergleich zu 1.2 um. Bei mittlerer Ausbildung wird die Wahrscheinlichkeit, als Arbeiterin erwerbstätig zu sein, beträchtlich vermindert. Die positiven Werte β_2^B, β_3^B bestätigen das bekannte Ergebnis, daß als Arbeiterinnen tätige Frauen im Durchschnitt eher Kinder haben als Angestellte. Je höher das Einkommen des Mannes ist, umso geringer wird die Wahrscheinlichkeit, als Arbeiterin beschäftigt zu sein.

6. Vergleich von GLM mit Goodman's ECTA und dem Regressionsansatz von Grizzle, Starmer und Koch (GSK)

6.1. Vergleich mit herkömmlichen loglinearen Modellen zur Analyse von Kontingenztabellen

Wie bereits gezeigt wurde, sind loglineare Modelle Spezialfälle verallgemeinerter linearer Modelle; alle mit ECTA oder einem der anderen Programme möglichen Analysen lassen sich daher auch mit GLIM durchführen. Die Vorteile von GLIM gegenüber ECTA liegen bei der Analyse mehrdimensionaler Kontingenztabellen vor allem im verwendeten Lösungsalgorithmus, ECTA verwendet zur Erzeugung der erwarteten Häufigkeiten unter einem loglinearen Modell den Deming-Stephan-Algorithmus der iterativen proportionalen Anpassung an im Modell vorgegebene Randverteilungen. Aus diesen geschätzten erwarteten Häufigkeiten werden - analog zur Varianzanalyse - die von uns mit $\underline{\beta}$ bezeichneten Regressionsparameter unter der Reparametrisierung mit zentrierten Effekten berechnet. Sobald nun fehlende Beobachtungen (strukturelle Nullen) auftreten, reduziert sich die Anzahl der Freiheitsgrade und die zentrierten Effekte lassen sich nicht mehr nach den gewohnten Formeln berechnen, da die Reparametrisierungen genau wie in der Varianzanalyse mit fehlenden Beobachtungen nicht mehr erfüllt werden können.

Dies kann an einem varianzanalytischen Beispiel leicht demonstriert werden.

Quantitative Beobachtungen y_{ij}, $i = 1,2$; $j = 1,2,3$ seien in der folgenden Tabelle zusammengefaßt.

	B_1	B_2	B_3
A_1	3	4	-
A_2	-	3	5

Nimmt man als saturiertes Modell

$$y_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}, \quad i = 1,2;\ j = 1,2,3$$

an und schätzt μ, α_i, β_j wie üblich nach der Formel

$$(6.1)\quad \hat{\mu} = \bar{\bar{y}} = \frac{1}{n} \sum_{i,j} y_{ij}$$

$$(6.2)\quad \hat{\alpha}_i = \bar{y}_{i.} - \bar{\bar{y}} \qquad i = 1,2$$

$$(6.3)\quad \hat{\beta}_j = \bar{y}_{.j} - \bar{\bar{y}} \qquad j = 1,2,3$$

$$(6.4)\quad \hat{\alpha\beta}_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{\bar{y}};\ i = 1,2;\ j = 1,2,3$$

So erhalten wir, wiederum in einer Tabelle zusammengefaßt

	β_1	β_2	β_3	
α_1	0,25	0,75	-	-0,25
α_2	-	-0,75	-0,25	0,25
	-0,75	-0,25	1,25	$\hat{\mu}$=3,75

Offensichtlich gelten folgende Reparametrisierungsbedingungen nicht

$$\sum_{j=1}^{3} \beta_j = \sum_{j=1}^{3} \alpha\beta_{ij} = \sum_{j=1}^{3} \alpha\beta_{2j} = \sum_{i=1}^{2} \alpha\beta_{i1} = \sum_{i=1}^{2} \alpha\beta_{i3} = 0.$$

Um trotzdem zu Schätzungen der Parameter zu kommen, müssen in ECTA Schätzwerte für fehlende Beobachtungen - in der Regel der Wert 0,5 - eingesetzt werden. Dies führt wiederum zu falschen Freiheitsgraden.

Im Gegensatz dazu wird im verallgemeinerten linearen Modell direkt nach den Parametern mit Hilfe des in 3. angegebenen iterativen Verfahrens aufgelöst. Die fehlenden Beobachtungen werden beim Aufbau der $\underline{X}$ Matrix der Dummyvariablen berücksichtigt, so daß wir sowohl die korrekte Anzahl der Freiheitsgrade als auch der Reparametrisierung entsprechende Schätzer der Regressionskoeffizienten erhalten.

Zusätzlich liefert uns der Kalkül eine Schätzung der Varianz-Kovarianzmatrix der Regressionskoeffizienten sowie die Möglichkeit, quantitiative Variable bei den unabhängigen Variablen zu berücksichtigen.

6.2. Vergleich mit dem GSK-Ansatz

Zur selben Zeit wie die loglinearen Modelle wurde als Alternative eine gewichtete kleinste Quadrate-Lösung zur Analyse von Kontingenztabellen von Grizzle, Starmer und Koch (1969) entwickelt und mit dem Programm NONMET II von Kritzer (1981) verbreitet. Eine kurze Darstellung der statistischen Grundlagen des GSK-Ansatzes und zahlreiche Beispiele sind in Küchler (1979) enthalten. Wir können uns daher beim Vergleich auf das Notwendigste beschränken.

Der GSK-Ansatz geht von folgender Oberlegung aus:

Gegeben sei eine qualitative abhängige Variable mit r Ausprägungen, die in s Subpopulationen, die durch hierarchische oder Kreuzklassifikationen definiert sind, auftritt. Ausgangspunkt ist eine Tabelle

$$(6.5)\quad \underline{N} = \begin{bmatrix} n_{11} & n_{12} \cdots & n_{1r} \\ \vdots & & \\ n_{j1} & n_{j2} \cdots & n_{jr} \\ \vdots & & \\ n_{s1} & n_{s2} \cdots & n_{sr} \end{bmatrix} \overset{\Sigma}{,} \begin{bmatrix} n_{1.} \\ \vdots \\ n_{2.} \\ \vdots \\ n_{s.} \end{bmatrix}$$

Wenn wir durch die Randsummen in den Subpopulationen dividieren, erhalten wir eine Matrix bedingter Prozentsätze.

$$(6.6)\quad \underline{P} = \begin{bmatrix} p_{11} & p_{12} \cdots & p_{1r} \\ \vdots & & \\ p_{j1} & p_{j2} \cdots & p_{jr} \\ \vdots & & \\ p_{s1} & p_{s2} \cdots & p_{sr} \end{bmatrix} \overset{\Sigma}{,} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Für jede Zeile von $\underline{P}$ ist die Varianz Kovarianz-Matrix der Prozentsätze, wenn π_{ji} der Erwartungswert von p_{ji} ist:

$$(6.7)\quad \underline{S}_j = \frac{1}{n_j} \begin{bmatrix} \pi_{j1}(1-\pi_{j1}) & -\pi_{j1}\pi_{j2} \cdots\cdots & \pi_{j1}\pi_{jr} \\ -\pi_{j2}\pi_{j1} & -\pi_{j2}(1-\pi_{j2}) \cdots & \pi_{j2}\pi_{jr} \\ \vdots & & \\ -\pi_{jr}\pi_{j1} & -\pi_{jr}\pi_{j2} \cdots\cdots & \pi_{jr}(1-\pi_{jr}) \end{bmatrix}$$

Wir fassen nun $\underline{P}$ als Spaltenvektor auf, indem wir die Zeilen transponieren und aufeinander stellen. Unter der Annahme der Unabhängigkeit der einzelnen Subpopulationen erhalten wir den Vektor $\underset{rsx1}{\underline{p}} = (p_k) \quad k = 1, \ldots\ r.s$

mit der Varianz Kovarianzmatrix

$$(6.8)\quad \underline{S} = \begin{bmatrix} \underline{S}_1 & & & 0 \\ & \underline{S}_2 & & \\ & & \underline{S}_j & \\ 0 & & & \underline{S}_s \end{bmatrix} \text{ mit } \underline{S}_j \text{ als Blockdiagonalelemente.}$$

Sind die Stichproben n_j genügend groß, also in etwa
$n_j\pi_{ji}(1-\pi_{ji}) \geq 9$, für alle $i = 1, \ldots\ r;\ j = 1, \ldots\ s$
ist der Vektor $\underline{p}$ in guter Näherung multivariat normalverteilt mit

$$(6.9)\quad \underline{p} \sim N(\underline{\pi}, \underline{S})$$

Da π_{ji} in $\underline{S}$ unbekannt ist, wird es durch p_{ji} geschätzt.

Wird nun an Stelle von $\underset{rs\times 1}{p}$ der Vektor der linearen Funktionen

$$(6.10)\quad \underset{m\times 1}{f} = \underset{m\times rs}{A}\ \underset{rs\times 1}{p}$$

betrachtet, so gilt exakt

$$(6.11)\quad \underline{f} \sim N(A\pi,\ \underbrace{ASA'}_{=V})$$

Ist $\underline{f}$ ein Vektor monotoner, zweimal differenzierbarer Funktionen von $\underline{p}$ mit $\underline{f} = \underline{f}(\underline{p})$ und ist $\underline{H} = \left(\frac{\partial f_l}{\partial p_{11}}\right)_{l,k}$ $l = 1,\dots m$; $k = 1,\dots rs$ die Matrix der ersten Ableitungen f_l nach p_k, so gilt wegen der Ergebnisse der Δ-Methode (Bishop et.al. 1975, S. 486 - 502) $\underline{f}$ ist angenähert normalverteilt

$$(6.12)\quad \underline{f} \sim N(\underline{f}(\pi),\ \underbrace{HSH'}_{=V})$$

Läßt sich nun $\underline{f}$ als lineare Funktion von unabhängigen Variablen darstellen, so gilt

$$(6.13)\quad \underline{f} \sim N(X\beta, V)$$

Wir erhalten einen Schätzwert für $\underline{\beta}$ indem wir eine gewichtete Regression durchführen. Wir erhalten als Lösung, wenn $Rg(X'V^{-1}X) = m$ ist:

$$(6.14)\quad \hat{\underline{\beta}} = (X'V^{-1}X)^{-1}X'V^{-1}f$$

mit

(6.15) $E\hat{\underline{\beta}} = \underline{\beta}$ und $\underline{V}_{\hat{\beta}} = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}$

sowie

(6.16) $\hat{\underline{\beta}} \sim N(\underline{\beta},(\underline{X}'\underline{V}^{-1}\underline{X})^{-1})$ asymptotisch

Bhapkar (1966) konnte nämlich zeigen, daß dieser Schätzer unter der Voraussetzung, daß alle $n_{ji} > 0$ sind, identisch ist mit Neyman's modifizierten Minimum Chi-Quadrat Schätzer, der zur Klasse der BAN-Schätzer (Best asymptotically normal) gehört.

Wie im verallgemeinerten linearen Modell lassen sich damit Konfidenzintervalle und Tests linearer Kontraste konstruieren.

Im Programm NONMET werden vor allem zwei Funktionstypen $\underline{f}$ verwendet.

Sei zunächst

$$\underline{f} = \underline{A}\underline{p} \quad \text{mit} \quad \underline{A} = \begin{bmatrix} \underline{A}_1 & & & 0 \\ & \underline{A}_2 & & \\ 0 & & \ddots & \\ & & & \underline{A}_s \end{bmatrix}$$

und $\underset{(r-1)xr}{\underline{A}_j} = \begin{bmatrix} 1 & 0 \ldots & 0 & 0 \\ 0 & 1 \ldots & 0 & 0 \\ \vdots & & & \vdots \\ 0 & \ldots\ldots & 1 & 0 \end{bmatrix}$, also eine lineare Funktion.

z. B. $r = 2$ $\quad \underline{A}_j = \begin{bmatrix} 1 & 0 \end{bmatrix}$

$r = 3$ $\quad \underline{A}_j = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

Im dichotomen Fall ist dann

$$(6.17)\quad \underline{P} = \begin{bmatrix} p_1 & 1-p_1 \\ p_2 & 1-p_2 \\ \vdots & \\ p_s & 1-p_s \end{bmatrix} \qquad \text{sowie}$$

$$(6.18)\quad \underline{f} = \underline{A}\underline{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_s \end{bmatrix}$$

$$(6.19)\quad \underline{V} = \operatorname{diag}\left(p_1(1-p_1)/n_1,\ p_2(1-p_2)/n_2,\ \ldots\ p_s(1-p_s)/n_s\right)$$

Die Lösung für $\hat{\underline{\beta}}$ entspricht dann einer gewichteten Regression mit Gewichten $w_j = \frac{1}{v_j} = \frac{n_j}{p_j(1-p_j)}$.

Wir erhalten exakt das gleiche Ergebnis, wenn wir im verallgemeinerten linearen Modell folgende Spezifikation vornehmen:

$$(6.20)\quad p_j = \pi_j + e_j \ ,$$

$$(6.21)\quad p_j \sim N(\pi_j, \frac{\pi_j(1-\pi_j)}{n_j})$$

$$(6.22)\quad \eta_j = \pi_j$$

$$(6.23)\quad \eta_j = \sum_{k=0}^{p} x_{jk}\beta_k$$

Als Gewicht w_j im Iterationsverfahren

$$w_j = \frac{1}{V_j}\left(\frac{d\pi_j}{d\eta_j}\right)^2$$

erhalten wir, da $\frac{d\pi_j}{d\eta_j} = 1$ ist.

$$w_j = \frac{1}{V_j} = \frac{n_j}{\pi_j(1-\pi_j)}\ ,\ \text{das durch}\ \frac{n_j}{p_j(1-p_j)}\ \text{geschätzt wird.}$$

Auf Grund der link Funktion $\eta_j = \pi_j$ ist nur ein Berechnungsschritt erforderlich.

Im dichotomen Fall ist der lineare GSK Ansatz sofort als Spezialfall von GLIM erkennbar.

Als Beispiel verwenden wir Daten von KÜchler (1981), die ausführlich mit NONMET analysiert wurden. Wir geben explizit die Designmatrix $\underline{X}$ mit den Spaltenvektoren C1, P1C1, PC13, GC2 an (konditionale Effekte), sowie die n_j, $j = 1,\ldots 18$ und die relativen Häufigkeiten p_j, $j = 1,\ldots 18$ an, um die Ergebnisse mit NONMET vergleichen zu können. Man beachte, daß für die

relative Häufigkeit 1.0 des 13. Elements im Vektor p der Wert (93-0.5)/93 eingesetzt wurde, um eine fehlende Beobachtung zu vermeiden. Dies wäre in GLIM nicht erforderlich. Die Ergebnisse für die Regressionskoeffizienten, die Teststatistiken und die Fehler sind - wie auf Grund von (6.20) - (6.23) zu erwarten war, genau gleich wie bei NONMET.

Die einzelnen GLIM Statements sind ebenfalls angegeben, um die Obersetzung von (6.20) - (6.23) in GLIM Sprache zu zeigen.

Tabelle 6.1.:

```
$C KUECHLERS DATEN MIT GLIM LINEARES MODELL MIT EINGELESENER DESIGNMATR
$UNITS 18
$DATA MEAN C1 P1C1 PC13 GC2 N PR $READ

1        1        0        1        0        21       0.80952
1        1        0        1        0        47       0.74468
1        0        1        0        1        39       0.38462
1        0        1        0        -1       110      0.10909
1        -1       0        1        0        30       0.1
1        -1       0        1        0        114      0.03509
1        1        0        0        0        72       0.91667
1        1        0        0        0        67       0.86567
1        0        0        0        1        25       0.56
1        0        0        0        -1       54       0.48148
1        -1       0        0        0        12       0.16667
1        -1       0        0        0        21       0.2381
1        1        0        -1       0        93       0.99465
1        1        0        -1       0        85       0.97647
1        0        -1       0        1        29       0.96552
1        0        -1       0        -1       27       0.88889
1        -1       0        -1       0        2        0.5
1        -1       0        -1       0        7        0.28571
$CALC V1 =1.0 - PR $CALC V = PR * V1 $CALC X = N / V
$YVAR PR $WEIGHT X
$ERR N
$FIT   C1 + P1C1 + PC13 + GC2
$DIS MERVS
 CYCLE  DEVIANCE      DF
    1    11.54        13

Y-VARIATE  PR
ERROR NORMAL  LINK IDENTITY
WEIGHT X
```

```
LINEAR PREDICTOR
<GM C1 P1C1 PC13 GC2

        ESTIMATE        S.E.        PARAMETER
     1    .5239        .7420E-02   <GM
     2    .3607        .1712E-01   C1
     3   -.3571        .2392E-01   P1C1
     4   -.1094        .1672E-01   PC13
     5    .6726E-01    .2255E-01   GC2
   SCALE PARAMETER TAKEN AS        .8881

   UNIT   OBSERVED      FITTED      RESIDUAL
      1     .8095        .7753        .3998
      2     .7447        .7753       -.4808
      3     .3046        .2341        1.932
      4     .1091        .9959E-01    .3198
      5     .1000E+00    .5388E-01    .8420
      6     .3509E-01    .5388E-01   -1.090
      7     .9167        .8846        .9839
      8     .8657        .8846       -.4549
      9     .5600        .5912       -.3141
     10     .4815        .4567        .3648
     11     .1667        .1632        .3190E-01
     12     .2381        .1632        .8055
     13     .9947        .9940        .8865E-01
     14     .9765        .9940       -1.065
     15     .9655        .9483        .5090
     16     .8889        .8138        1.242
     17     .5000        .2726        .6432
     18     .2857        .2726        .7680E-01

(CO)VARIANCE MATRIX
   1  5.5056E-05
   2 -4.1764E-05  2.9326E-04
   3  4.3411E-06 -3.2930E-06  5.7197E-04
   4 -3.0737E-06  2.5511E-04 -2.4236E-07  2.7954E-04
   5  1.2124E-05 -9.1969E-06  3.2517E-04 -6.7686E-07  5.0846E-04
        1           2           3           4           5
    SCALE PARAMETER TAKEN AS     .8881

S.E. OF DIFFERENCES
   1  0.
   2  2.0781E-02  0.
   3  2.4866E-02  2.9526E-02  0.
   4  1.8459E-02  7.9109E-03  2.9189E-02  0.
   5  2.3222E-02  2.8638E-02  2.0738E-02  2.8095E-02  0.
        1           2           3           4           5
```

Im trichotomen Fall erhalten wir

$$(6.24)\quad \underline{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ \vdots & & \\ p_{s1} & p_{s2} & p_{s3} \end{bmatrix} \qquad \text{sowie}$$

$$(6.25)\quad \underline{f} = \underline{A}\underline{p} = \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \\ p_{22} \\ \vdots \\ p_{j1} \\ p_{j2} \\ \vdots \\ p_{s1} \\ p_{s2} \end{bmatrix}$$

$$(6.26)\quad \underline{V} = \begin{bmatrix} \underline{V}_1 & & & 0 \\ & \underline{V}_2 & & \\ & & \ddots & \\ & & \underline{V}_j & \\ 0 & & & \ddots \\ & & & & \underline{V}_s \end{bmatrix} \quad \text{mit } \underline{V}_j = \frac{1}{n_j} \begin{bmatrix} p_{j1}(1-p_{j1}) & -p_{j1}p_{j2} \\ -p_{j2}p_{j1} & p_{j2}(1-p_{j2}) \end{bmatrix}^{-1}$$

Dieses Modell läßt sich nicht unmittelbar in verallgemeinerte lineare Modelle einbetten, da wir im Iterationsverfahren nur Diagonalmatrizen als Gewichte verwendet haben.

Trotzdem läßt sich auch dieser Fall in das Programmsystem GLIM einbetten, indem wir mit Hilfe eines GLIM Unterprogramms die Choleskymatrix $\underline{V}^{\frac{1}{2}}$ von $\underline{V}$ und ihre Inverse berechnen und die übliche Regression mit den transformierten Variablen

$$(6.27)\quad \underline{q} = \underline{V}^{-\frac{1}{2}}\underline{p},\ \underline{Z} = \underline{V}^{-\frac{1}{2}}\underline{X}$$

durchführen.

$$(6.28)\quad \hat{\underline{\beta}} = (\underline{Z}'\underline{Z})^{-1}\underline{Z}'\underline{q} = (\underline{X}'\underline{V}^{-\frac{1}{2}'}\underline{V}^{-\frac{1}{2}}\underline{X})^{-1}\underline{X}'\underline{V}^{-\frac{1}{2}'}\underline{V}^{-\frac{1}{2}}\underline{q}$$

$$= (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{q}$$

Als zweiter Funktionstyp wird in NONMET die logit Transformation verwendet.

Wiederum untersuchen wir zunächst den dichotomen Fall.

$$(6.29)\quad \underline{P} = \begin{bmatrix} p_1 & (1-p_1) \\ p_2 & (1-p_2) \\ \vdots & \\ p_j & (1-p_j) \\ \vdots & \\ p_s & (1-p_s) \end{bmatrix}.$$

$$(6.30)\quad f_j = \ln \frac{p_j}{1-p_j} \qquad j = 1, \ldots s$$

$$(6.31)\quad \underline{H} = \left(\frac{\partial f_j}{\partial p_l}\right)_{j,l} \qquad j = 1,\dots s;\ l = 1,\dots s$$

$$\left(\frac{\partial f_j}{\partial p_l}\right) = \begin{cases} \dfrac{1}{p_j(1-p_j)} & l = j \\ 0 & l \neq j \end{cases}$$

Da $\underline{S} = \operatorname{diag}\left(\frac{p_1(1-p_1)}{n_1} \dots \frac{p_j(1-p_j)}{n_j} \dots \frac{p_s(1-p_s)}{n_s}\right)$, gilt

$$(6.32)\quad \underline{V} = \underline{HSH}' = \operatorname{diag}\left(\frac{1}{p_1(1-p_1)n_1} \dots \frac{1}{p_j(1-p_j)n_j} \dots \frac{1}{p_s(1-p_s)n_s}\right)$$

$$(6.33)\quad \underline{V}^{-1} = \operatorname{diag}\left(n_1p_1(1-p_1) \dots n_jp_j(1-p_j) \dots n_sp_s(1-p_s)\right)$$

Wir vergleichen diese gewichtete Regression mit dem Iterationsverfahren im verallgemeinerten linearen Modell

$$(6.34)\quad y_j = \pi_j + e_j$$

y_j sei binomialverteilt mit $Ey_j = \pi_j$ und $V(y_j) = \frac{\pi_j(1-\pi_j)}{n_j}$

$$(6.35)\quad n_j = \ln \frac{\pi_j}{1-\pi_j}$$

$$(6.36)\quad n_j = \sum_{k=0}^{p} x_{jk}\beta_k$$

Die Gewichte in der iterierten gewichteten Regression werden berechnet (vgl. Gleichung(3.6))

$$(6.37)\quad w_j^q = \frac{1}{V(y_j^q)} \left(\frac{d\pi_j^q}{dn_j^q}\right)^2 \qquad \text{mit } q = \text{Iterationszahl}$$

Wie bereits in 3.1 als Spezialfall von (3.6) gezeigt, erhalten wir

$$(6.38)\quad w_j^q = \frac{n_j}{\pi_j^q(1-\pi_j^q)} \quad (\pi_j^q(1-\pi_j^q))^2$$

und damit für die Matrix der Gewichte

$$\underline{W}^q = \text{diag}\{n_1\pi_1^q(1-\pi_1^q)\ldots\ n_j\pi_j^q(1-\pi_j^q)\ldots\ n_s\pi_s^q(1-\pi_s^q)\}$$

Im Iterationsverfahren wird als Startwert $\pi_j^o = p_j$ gesetzt. $\underline{W}^o$ ist daher identisch mit $\underline{V}^{-1}$ aus dem GSK Ansatz.

Der GSK Ansatz liefert daher für den dichotomen Fall die Werte des ersten Iterationsschrittes im verallgemeinerten linearen Modell.

Als Default Annahme in NONMET wird für den polytomen Fall eine Auflösung in logits in folgender Form vorgenommen:

$$(6.39)\quad f_{j1} = \ln\frac{p_{j1}}{1-p_{j1}}$$

$$f_{j2} = \ln\frac{p_{j2}}{1-p_{j2}}$$

$$\vdots$$

$$f_{jr-1} = \ln\frac{p_{jr-1}}{1-p_{jr-1}}$$

Für f_{j1}, f_{j2} ... werden getrennte logit Analysen vorgenommen, so daß wir auf den dichotomen Fall zurückverweisen können.

Selbstverständlich läßt sich der GSK Ansatz auch auf andere Funktionstypen erweitern, die für spezielle Fragestellungen zugeschnitten sind und erweist sich damit als flexibel in den Anwendungen. Für Anwendungen ist dieser Ansatz dann problematisch, wenn leere Zellen auftreten und/oder die Randsummen n_j relativ klein sind. Die in NONMET vorgesehene Regelung, fehlende Zellen mit 1/r zu besetzen, birgt dieselben Probleme, die bereits bei ECTA besprochen wurden, in sich.

Statistisch gesehen, sind zwar auch die GSK Schätzer BAN Schätzer, erfüllen aber nicht wie die ML Schätzer die Bedingungen der Wirksamkeit zweiter Ordnung (Rao (1962)). Die Varianzen der ML Schätzer sind bei kleinen Stichproben kleiner als bei anderen Schätzverfahren. Weiter konnte Habermann (1977) beweisen, daß für die ML Schätzer bei loglinearen Modellen schwächere Bedingungen für asymptotische Resultate genügen als in der üblichen ML Theorie.

Obwohl es leicht durchführbar wäre, wurde bisher in NONMET nicht die Möglichkeit eingebaut, auch quantitative Variable als unabhängige Variable zuzulassen.

7. Erweiterungen

Neben den bisherigen "klassischen" Anwendungen des verallgemeinerten linearen Modells lassen sich auch folgende, den Sozialwissenschaftler interessierende und in letzter Zeit mehrfach diskutierte Modelle auf das allgemeine Modell zurückzuführen.

- die Behandlung von ordinalen abhängigen Variablen

- die Analyse von Übergangsraten inhomogener Markoffprozesse, die bei life event histories von wesentlicher analytischer Bedeutung sind

- die Maximum Likelihood Schätzung von latent structure Modellen

Die Erweiterungen werden in Arminger (1982) durchgeführt.

A. Mathematischer Anhang

A.1 Beweis der Gleichungen (2.5) und (2.6)

Wir gehen von der log likelihood Funktion

$L(\theta) = \ln p(y)$ aus .

$$L(\theta) = \{[y\theta - b(\theta)]/a(\phi) + c(y,\phi)\}$$

Wir verwenden die allgemein für log likelihood Funktionen unter Regularitätsvoraussetzungen bewiesenen Eigenschaften (vgl. Kendall and Stuart (1979), S. 9).

$$(A1)\quad E\left(\frac{\partial L}{\partial \theta}\right) = 0$$

$$(A2)\quad E\left(\frac{\partial^2 L}{\partial \theta^2}\right) + E\left(\left(\frac{\partial L}{\partial \theta}\right)^2\right) = 0$$

Zunächst gilt:

$$\frac{\partial L}{\partial \theta} = \frac{1}{a(\phi)}\,(y - b'(\theta))$$

$$0 = E\left(\frac{\partial L}{\partial \theta}\right) = \frac{1}{a(\phi)}\,[E(y) - b'(\theta)] \quad \Longrightarrow$$

$$(A3)\quad E(y) = \mu = b'(\theta).$$

Daraus folgt

$$V(y) = E(y-\mu)^2 = E(y-b'(\theta))^2$$

$$E\left(\frac{\partial L}{\partial \theta}\right)^2 = \frac{1}{a(\phi)^2} E[(y-b'(\theta))^2] = \frac{1}{a(\phi)^2} V(y)$$

$$E\left(\frac{\partial^2 L}{\partial \theta^2}\right) = \frac{-b''(\theta)}{a(\phi)}$$

Wegen (A2) gilt nun

(A4) $V(y) = b''(\theta)a(\phi)$

A.2 Multinomial- und Poissonverteilung

(A5) Satz: Die Dichte eines multinomial verteilten Zufallsvektors $(X_1 \dots X_k)$ mit $\sum_{l=1}^{k} X_l = N$ läßt sich als Produkt von k Dichten unabhängiger poissonverteilter Zufallsvariablen darstellen, gegeben $\sum_{l=1}^{k} X_l = N$ ist ebenfalls poissonverteilt.

Beweis:

$$p(X_1=x_1,\ X_2=x_2 \dots X_k=x_k) = \frac{N!}{x_1!x_2! \dots x_k!}\, \pi_1^{x_1} \pi_2^{x_2} \dots \pi_k^{x_k}$$

mit π_l als Wahrscheinlichkeit, daß Kategorie l angenommen wird, ist Dichte der Multinomialverteilung.

Seien nun $\mu_1 \dots \mu_k$ die Intensitätsparameter der poissonverteilten Zufallsvariablen.

Dann gilt für die poissonverteilten Zufallsvariablen

$$P(X_1=x_1,\ X_2=x_2 \dots X_k=x_k / \sum_{l=1}^{k} X_l = N)$$

$$= \frac{\mu_1^{x_1}}{x_1!} e^{-\mu_1} \frac{\mu_2^{x_2}}{x_2!} e^{-\mu_2} \dots \frac{\mu_k^{x_k}}{x_k!} e^{-\mu_k} \Big/ \frac{(\sum_{l=1}^{k} \mu)^N}{N!} e^{-\sum_{l=1}^{k} \mu_l}$$

$$= \frac{N!}{x_1!x_2! \dots x_k!} \left(\frac{\mu_1}{\Sigma\mu_l}\right)^{x_1} \dots \left(\frac{\mu_k}{\Sigma\mu_l}\right)^{x_k}$$

Wenn wir nun $\pi_i = \left(\frac{\mu_j}{\Sigma\mu_l}\right)$ setzen, erhalten wir den Behauptung.

A.3 Beweis des Schätzverfahrens

Die Maximum-Likelihood-Methode, deren Eigenschaften etwa bei Kendall and Stuart (1979, S. 45) ausführlich beschrieben werden, liefert uns unter schwachen Bedingungen asymptotisch erwartungstreue, normalverteilte Schätzer von $\underline{\beta}$ mit kleinstmöglicher Varianz.

Zu diesem Zweck maximieren wir die log likelihood Funktion

$$(A6)\quad L = \sum_{i=1}^{n} L_i \quad \text{mit} \quad L_i = [y_i \theta_i - b(\theta_i)]/a_i(\phi) + c(y_i, \phi)$$

in Abhängigkeit von β_j, $j = 1, \ldots p$, indem wir den Vektor der ersten Ableitungen $\frac{\partial L}{\partial \underline{\beta}}$, die Matrix der zweiten Ableitungen $\frac{\partial^2 L}{\partial \underline{\beta} \partial \underline{\beta}}$ bilden und $\frac{\partial L}{\partial \underline{\beta}} = 0$ setzen.

Wegen $L = \sum_{i=1}^{n} L_i$ gilt $\frac{\partial L}{\partial \underline{\beta}} = \sum_{i=1}^{n} \frac{\partial L_i}{\partial \underline{\beta}}$. Weiter ist

$$\frac{\partial L_i}{\partial \underline{\beta}} = \left(\frac{\partial L_i}{\partial \beta_j}\right)_{j=1,\ldots p} \quad \text{und}$$

$$(A7)\quad \frac{\partial L_i}{\partial \beta_j} = \frac{dL_i}{d\theta_i} \cdot \frac{d\theta_i}{d\mu_i} \cdot \frac{d\mu_i}{d\eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}$$

Wir bilden nun die einzelnen Ableitungen

$$(A8)\quad \frac{dL_i}{d\theta_i} = (y_i - b'(\theta_i))/a_i(\phi) \underset{(A3)}{=} (y_i - \mu_i)/a_i(\phi)$$

$$(A9)\quad \frac{d\theta_i}{d\mu_i} = \left(\frac{d\mu_i}{d\theta_i}\right)^{-1} \underset{(A3)}{=} (b''(\theta_i))^{-1} \underset{(A4)}{=} \frac{a_i(\phi)}{V(y_i)}$$

$$(A10)\quad \frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$$

Damit gilt insgesamt

$$(A11)\quad \frac{\partial L_i}{\partial \beta_j} = \left(\frac{y_i-\mu_i}{V(y_i)}\right) \cdot \frac{d\mu_i}{d\eta_i} x_{ij} = 0 \qquad \text{und}$$

$$(A12)\quad \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{y_i-\mu_i}{V(y_i)}\right) \frac{d\mu_i}{d\eta_i} x_{ij} = 0 \quad j = 1,\dots p$$

Wenn

$$\theta_i = \eta_i$$

so gilt

$$\frac{d\mu_i}{d\eta_i} = \frac{d\mu_i}{d\theta_i}$$

und wir erhalten für

$$(A13)\quad \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \frac{dL_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\theta_i} \frac{\partial \theta_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{dL_i}{d\theta_i} \frac{\partial \theta_i}{\partial \beta_j}$$

$$= \sum_{i=1}^{n} \left(\frac{y_i-\mu_i}{a_i(\phi)}\right) x_{ij} = 0$$

Daraus folgt unmittelbar, daß

$$(A14)\quad \sum_{i=1}^{n} y_i x_{ij} = \sum_{i=1}^{n} \mu_i x_{ij}$$

bei der ML-Lösung sein muß.

Wenn die x_{ij} Dummy-Variable sind, erhalten wir für den Fall der poissonverteilten Zufallsvariablen die aus den log linearen Modellen wohlbekannte Forderung, daß die Randverteilungen der beobachteten Werte $(\sum_{i=1}^{n} y_i x_{ij})$ mit den Randverteilungen der erwarteten Werte $(\sum_{i=1}^{n} \mu_i x_{ij})$ übereinstimmen müssen.

Weiter läßt sich für den Fall $\theta_i = n_i$ zeigen, daß $\sum_{i=1}^{n} \frac{1}{a_i(\phi)} y_i x_{ij}$ suffiziente Statistiken der Parameter β_j, $j = 1,\dots p$ sind. Es gilt nämlich für $\theta_i = n_i = \sum_{j=1}^{p} \beta_j x_{ij}$,

$$L = \sum_{i=1}^{n} \left([y_i \theta_i - b(\theta_i)]/a_i(\phi) + c(y_i, \phi) \right)$$

$$= \sum_{j=1}^{p} \beta_j \sum_{i=1}^{n} \frac{1}{a_i(\phi)} y_i x_{ij} - \sum_{i=1}^{n} \frac{1}{a_i(\phi)} b(n_i) + \sum_{i=1}^{n} c(y_i, \phi)$$

wodurch das Faktorisierungskriterium erfüllt ist.

Die Matrix der zweiten Ableitungen ergibt sich aus

$$\text{(A15)} \quad \frac{\partial^2 L}{\partial \beta_j \partial \beta_k} = \frac{\partial}{\partial \beta_k} \left(\sum_{i=1}^{n} \left(\frac{y_i - \mu_i}{V(y_i)} \right) \frac{d\mu_i}{dn_i} x_{ij} \right)$$

Da $V(y_i) = b''(\theta_i) a_i(\phi)$ gemäß (A4), müssen sowohl μ_i als auch $V(y_i)$ partiell nach β_k abgeleitet werden. Für das einzelne Element $j,k = 1,\dots p$ gilt:

$$(A16)\quad \frac{\partial^2 L}{\partial\beta_j \partial\beta_k} = \sum_{i=1}^{n} \frac{-1}{V(y_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{ik} + \qquad (I)$$

$$+ \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial V(y_i)^{-1}}{\partial\beta_k} \frac{d\mu_i}{d\eta_i} x_{ij} + \qquad (II)$$

$$+ \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(y_i)} \frac{d^2\mu_i}{d\eta_i^2} x_{ij}x_{ik} \qquad (III)$$

Bezeichnen wir mit G die Matrix der zweiten Ableitungen

$$(g_{ik})_{j,k=1,\ldots p} = \left(\frac{\partial^2 L}{\partial\beta_j \partial\beta_k}\right)_{j,k=1,\ldots p}$$

erhalten wir für den Erwartungswert

$$E(G) = E\left(\frac{\partial^2 L}{\partial\beta_j \partial\beta_k}\right)_{j,k=1,\ldots p}$$

da $Ey_i = \mu_i$ ist

$$(A17)\quad h_{jk} = E(g_{jk}) = \left(\sum_{i=1}^{n} \frac{-1}{V(y_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ik}x_{ij}\right)_{j,k=1,\ldots p}$$

Ist wiederum $\theta_i = \eta_i$, so gilt

$$(A18)\quad \underline{H} = \underline{G}$$

Die Matrix der zweiten Ableitungen ist gleich ihrem Erwartungswert.

Beweis: Wir zeigen, daß der Summand II = -III in (A16), wenn $\theta_i = \eta_i$.

$$\sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial V(y_i)^{-1}}{\partial \beta_k} \frac{d\mu_i}{d\eta_i} x_{ij} =$$

$$= \sum_{i=1}^{n} (y_i - \mu_i) (-1) V(y_i)^{-2} \frac{dV(y_i)}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_k} \frac{d\mu_i}{d\eta_i} x_{ij} =,$$

$$\text{da } V(y_i) = b''(\eta_i) a_i(\phi) = \frac{d\mu_i}{d\eta_i} a_i(\phi),$$

$$= \sum_{i=1}^{n} (y_i - \mu_i) (-1) V(y_i)^{-2} \frac{d^2\mu_i}{d\eta_i^2} a_i(\phi) x_{ik} \frac{d\mu_i}{d\eta_i} x_{ij}$$

$$= \sum_{i=1}^{n} (y_i - \mu_i) (-1) V(y_i)^{-2} \frac{d^2\mu_i}{d\eta_i^2} x_{ik} V(y_i) x_{ij}$$

$$= -1 \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(y_i)} \frac{d^2\mu_i}{d\eta_i^2} x_{ij} x_{ik} = -III.$$

Sei nun $\underset{n\times n}{W}$ eine Diagonalmatrix von Gewichten $\underset{n\times n}{W} = \text{diag}\{w_i\}$ mit

$$(A19)\quad w_i = \frac{1}{V(y_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2$$

so können wir das Newtonverfahren zur Berechnung der Regressionskoeffizienten in folgender Weise ansetzen (zum Newtonverfahren vgl. z. B. Luenberger (1973)).

Sei q der Laufindex

FÜr q = 0 wählen wir als Startwerte fÜr $\underline{\mu}$ die Beobachtungswerte $\underline{y}$.

Sei $\underline{z}^q$ der Vektor der ersten Ableitungen der log likelihood Funktion nach $\underline{\beta}$ an der Stelle $\underline{b}^q$ und $\underline{W}^q$ die Diagonalmatrix der Gewichte.

$$z_j^{*q} = \sum_{i=1}^{n} \frac{1}{V^q(y_i)} (y_i - \mu_i^q) \left(\frac{d\mu_i^q}{d\eta_i^q}\right) x_{ij} \qquad j = 1, \ldots p$$

Dann gilt bei Fisher's scoring Methode, wenn wir bei der Newton-Methode an Stelle von g_{jk} den Erwartungswert h_{jk} nehmen

$$(A20) \quad -\underline{H}^q(\underline{b}^{q+1} - \underline{b}^q) = \underline{z}^{*q}$$

In Vektor- bzw. Matrixschreibweise gilt mit den obigen Bezeichnungen

$$(A21) \quad -\underline{H}^q = \underline{X}'\underline{W}^q\underline{X}$$

$$(A22) \quad \underline{z}^{*q} = \underline{X}'\underline{W}^q\underline{r}^q \quad \text{mit}$$

$$(A23) \quad r_i^q = (y_i - \mu_i^q) \frac{d\eta_i^q}{d\mu_i^q}$$

so daß wir erhalten

$$(A24) \quad \underline{X}'\underline{W}^q\underline{X}(\underline{b}^{q+1} - \underline{b}^q) = \underline{X}'\underline{W}^q\underline{r}^q$$

Wir formen um:

$$(\underline{X}'\underline{W}^q\underline{X})\underline{b}^{q+1} = \underline{X}'\underline{W}^q(\underbrace{\underline{X}\underline{b}^q+\underline{r}^q}_{\underline{n}^q})$$

$$(A25)\quad (\underline{X}'\underline{W}^q\underline{X})\underline{b}^{q+1} = \underline{X}'\underline{W}^q(\underline{n}^q+\underline{r}^q)$$

Daraus folgt - sofern $\underline{X}$ von vollem Spaltenrang ist - die Lösung

$$(A26)\quad \underline{b}^{q+1} = (\underline{X}'\underline{W}^q\underline{X})^{-1}\underline{X}'\underline{W}^q(\underline{n}^q+\underline{r}^q)$$

Dies entspricht einer gewichteten Regression mit Gewichtsmatrix $\underline{W}$ und abhängigen Variablen $\underline{n}^q+\underline{r}^q$.

Der Iterationsprozeß wird solange wiederholt , bis

$$(A27)\quad \sum_{j=1}^{p} |\beta_j^{q+1}-\beta_j^q| < \varepsilon \text{ , z. B } \varepsilon = 0{,}0001$$

Wenn wir den nach Abschluß des Iterationsprozesses erhaltenen Regressionsvektor mit $\underline{b}$, die Gewichtsmatrix mit $\underline{W}$, die geschätzten Werte von $\underline{n}$ mit $\hat{\underline{n}}$ und die von $\underline{\mu}$ mit $\hat{\underline{\mu}}$ bezeichnen, so wird die Fisher'sche Informationsmatrix von $\hat{\underline{\beta}}$ durch $\underline{X}'\underline{W}\underline{X}$ geschätzt und ihre Inverse ist ein Schätzer für die Varianz-Kovarianzmatrix von $\underline{b}$. (vgl. Kendall and Stuart (1979), S. 59) Auf Grund allgemeiner Ergebnisse der Maximum-Likelihood-Schätzung ist dann der Regressionsvektor $\underline{b}$ asymptotisch normalverteilt mit Erwartungswert $\underline{\beta}$ und Varianz-Kovarianzmatrix $(\underline{X}'\underline{W}\underline{X})^{-1}$.

A.4 Likelihood-Ratio-Test

Sei $\tilde{L}_c$ die Likelihood-Funktion des laufenden Modells mit kanonischen Parametern ${}_c\theta_i$ und geschätzten Erwartungswerten ${}_c\mu_i$. Analog seien $\tilde{L}_f$, ${}_f\theta_i$ und ${}_f\mu_i = y_i$ im saturierten Modell definiert.

L_c und L_f seien die log Likelihood-Funktionen. Bekanntlich gilt unter schwachen Bedingungen (vgl. Kendall and Stuart (1979), S. 246)

(A28) $S(c,f) = -2 \ln \frac{\tilde{L}_c}{\tilde{L}_f}$ ist unter der H_0: $\beta_{p+1} = \beta_{p+2} = \ldots \beta_n = 0$

χ^2 verteilt mit n-p Freiheitsgraden.

Da $S(c,f) = -2 \ln \frac{\tilde{L}_c}{\tilde{L}_f} = 2(L_f - L_c)$ ist

$$\frac{1}{2} S(c,f) = \sum_{i=1}^{n} \frac{1}{a_i(\phi)} (y_i \, {}_f\theta_i - b({}_f\theta_i)) + c(y_i,\phi)$$

$$- \sum_{i=1}^{n} \frac{1}{a_i(\phi)} (y_i \, {}_c\theta_i - b({}_c\theta_i)) + c(y_i,\phi)$$

$$= \sum_{i=1}^{n} \frac{1}{a_i(\phi)} (y_i({}_f\theta_i - {}_c\theta_i) + b({}_c\theta_i) - b({}_f\theta_i))$$

(A29) $$S(c,f) = 2 \sum_{i=1}^{n} \frac{1}{a_i(\phi)} (y_i({}_f\theta_i - {}_c\theta_i) + b({}_c\theta_i) - b({}_f\theta_i))$$

Unter Verwendung des obigen Satzes läßt sich für hierarchische Hypothesen folgende Aussage treffen:

Sei M_{c_1} ein Modell mit Parametern $\beta_1, \ldots \beta_k \neq 0$, $k < p$
und $\tilde{L}_{c_1}$ die zugehörige Likelihoodfunktion.

Sei M_{c_2} ein hierarchisch übergeordnetes Modell mit Parametern $\beta_1, \ldots \beta_k, \beta_{k+1} \ldots \beta_p \neq 0$ und $\tilde{L}_{c_2}$ die zugehörige Likelihoodfunktion.

$\tilde{L}_f$ sei wie oben definiert. Dann gilt

$$(A30)\quad S(c_1,c_2) = -2 \ln \frac{\tilde{L}_{c_1}}{\tilde{L}_{c_2}} = -2 \ln \frac{\tilde{L}_{c_1}/\tilde{L}_f}{\tilde{L}_{c_2}/\tilde{L}_f} = S(c_1,f) - S(c_2,f)$$

ist unter der Hypothese H_0: $\beta_{k+1} \ldots \beta_n = 0$ asymptotisch χ^2 verteilt mit p-k Freiheitsgraden.

A.5 Die Multinomialverteilung als Produkt von Binomialverteilungen

Die Zufallsvariablen $X_1 \ldots X_K$ seien multinomialverteilt mit

$$(A31)\quad P(X_1 = x_1,\ X_2 = x_2 \ldots X_k = x_k / \Sigma x_i = n) = \frac{n!}{x_1! x_2! \ldots x_K!}\, \pi_1^{x_1} \pi_2^{x_2} \ldots \pi_K^{x_K}$$

Wir definieren die binomialverteilten Zufallsvariablen X_j, $j = 1, \ldots K-1$ mit

$$(A32)\quad P(X_j = x_j) = \binom{n_j}{x_j} \lambda_j^{x_j} (1-\lambda_j)^{n_j - x_j} \qquad j = 1, \ldots K-1$$

mit $n_j = n \qquad j = 1$

$$n_j = n - \sum_{l=1}^{j-1} x_l, \qquad j = 2 \ldots K-1$$

$$\lambda_j = \pi_j / \left(\sum_{l=j}^{K} \pi_l \right)$$

Dann gilt

$$(A33)\quad \prod_{j=1}^{K-1} P(X_j = x_j) = \frac{n!(n-x_1)!(n-x_1-x_2)! \ldots \left(n - \sum_{l=1}^{K-2} x_l\right)!}{x_1!(n-x_1)!x_2!(n-x_1-x_2)! \ldots x_{K-1}! x_K!} \cdot$$

$$\cdot\ \pi_1^{x_1} \left(\sum_{l=2}^{K} \pi_2 \right)^{n-x_1} \left(\frac{\pi_2}{\sum_{l=2}^{K} \pi_l} \right)^{x_2} \left(\sum_{l=3}^{K} \pi_l \Big/ \sum_{m=2}^{K} \pi_m \right)^{n-x_1-x_2} \ldots \left(\frac{\pi_{K-1}}{\pi_{K-1}+\pi_K} \right)^{x_{K-1}} \left(\frac{\pi_K}{\pi_{K-1}+\pi_K} \right)^{x_K}$$

Kürzen und Zusammenfassen ergibt dann die in (A31) angeführten Likelihoodfunktion der Multinomialverteilung.

B. Anhang: probit und komplementäre log log link Funktionen für das einführende Beispiel

Y-VARIATE R
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR N

LINEAR PREDICTOR
<GM A B Z A.B B.Z

	ESTIMATE	S.E.	PARAMETER
1	-.7392	.7192E-01	<GM
2	-.3850	.4706E-01	A(2)
3	-.1154	.1677	A(3)
4	.3910	.8245E-01	B(2)
5	1.074	.8754E-01	B(3)
6	.4719E-01	.6348E-02	Z
7	.3678	.5282E-01	A(2).B(2)
8	.2811	.5560E-01	A(2).B(3)
9	.1017	.2090	A(3).B(2)
10	.8336E-01	.2002	A(3).B(3)
11	.2697E-01	.7299E-02	B(2).Z
12	.1904E-01	.7714E-02	B(3).Z

SCALE PARAMETER TAKEN AS 1.000

UNIT	OBSERVED	OUT OF	FITTED	RESIDUAL
1	16	32	9.572	2.482
2	52	96	47.44	.9302
3	43	57	41.98	.3064
4	5	16	2.895	1.367
5	13	35	17.06	-1.372
6	17	26	18.24	-.5308
7	132	383	130.7	.1383
8	640	1155	655.8	-.9407
9	607	793	624.6	-1.526
10	47	217	46.35	.1070
11	260	461	258.6	.1279
12	265	364	275.3	-1.254
13	1	3	.9001	.1259
14	0	1	.7782	-1.873
15	329	845	333.4	-.3118
16	2925	4398	2872.	1.689
17	2838	3359	2824.	.6831
18	242	913	234.7	.5521
19	1874	2926	1892.	-.6896
20	2384	2877	2342.	2.012
21	1	13	4.562	-2.070
22	9	14	9.070	-.3931E-01
23	13	15	12.49	.3527
24	100	207	100.9	-.1257
25	927	1246	968.9	-2.852
26	1022	1126	1022.	-.3088E-01
27	178	617	208.9	-2.627
28	1581	2036	1573.	.4426
29	2118	2420	2153.	-2.247
30	10	23	10.16	-.6604E-01
31	39	51	44.31	-2.204
32	95	109	98.37	-1.089
33	16	32	18.59	-.9288
34	143	162	141.2	.4166
35	147	153	145.6	.5373
36	106	199	85.24	2.973
37	722	820	711.9	1.046
38	908	960	902.4	.7558
39	16	23	12.32	1.540
40	94	102	88.63	1.577
41	209	217	205.7	.9965

Y-VARIATE R
ERROR BINOMIAL LINK C-LOGLOG
BINOMIAL DENOMINATOR N

LINEAR PREDICTOR
<GM A B Z A.B B.Z

	ESTIMATE	S.E.	PARAMETER
1	-1.331	.9476E-01	<GM
2	-.5104	.6257E-01	A(2)
3	-.1396	.2006	A(3)
4	.6854	.1023	B(2)
5	1.408	.1030	B(3)
6	.6327E-01	.8175E-02	Z
7	.4969	.6683E-01	A(2).B(2)
8	.4235	.6715E-01	A(2).B(3)
9	.1042	.2257	A(3).B(2)
10	.1076	.2163	A(3).B(3)
11	.6374E-02	.8818E-02	B(2).Z
12	-.1040E-01	.8868E-02	B(3).Z

SCALE PARAMETER TAKEN AS 1.000

UNIT	OBSERVED	OUT OF	FITTED	RESIDUAL
1	16	32	9.475	2.526
2	52	96	49.14	.5845
3	43	57	42.51	.1492
4	5	16	3.041	1.249
5	13	35	17.75	-1.606
6	17	26	18.59	-.6918
7	132	383	129.1	.3081
8	640	1155	663.1	-1.373
9	607	793	626.9	-1.735
10	47	217	47.47	-.7767E-01
11	260	461	262.4	-.2256
12	265	364	277.2	-1.494
13	1	3	.9021	.1232
14	0	1	.7799	-1.883
15	329	845	331.1	-.1460
16	2925	4398	2862.	1.999
17	2838	3359	2821.	.7936
18	242	913	235.6	.4834
19	1874	2926	1889.	-.5981
20	2384	2877	2340.	2.086
21	1	13	4.564	-2.071
22	9	14	8.928	.4023E-01
23	13	15	12.46	.3748
24	100	207	102.4	-.3293
25	927	1246	965.2	-2.588
26	1022	1126	1022.	-.4334E-01
27	178	617	207.3	-2.501
28	1581	2036	1568.	.6909
29	2118	2420	2148.	-1.962
30	10	23	10.29	-.1230
31	39	51	44.35	-2.225
32	95	109	98.19	-1.022
33	16	32	19.45	-1.250
34	143	162	142.4	.1516
35	147	153	146.2	.3274
36	106	199	85.55	2.928
37	722	820	717.8	.4445
38	908	960	904.4	.4928
39	16	23	12.81	1.339
40	94	102	88.70	1.559
41	209	217	206.3	.8453

Literaturverzeichnis

Anderson, T.W. (1958), An introduction to multivariate statistical analysis.

Arminger, G. (1976), Loglineare Modelle zur Analyse nominal skalierter Variablen, Wien.

Arminger, G. (1979), Loglineare Modelle zur Analyse des Zusammenhangs zwischen nominalen Variablen, in Die Befragung 6, Hrsg. K. Holm, S. 218 - 261, München.

Arminger, G. (1979), Das allgemeine logistische Modell, in Die Befragung 6,Hrsg. K. Holm, S. 262 - 272, München.

Arminger, G., Lijphart, N., Müller, W. (1981), Die Verwendung log-linearer Modelle zur Disaggregierung aggregierter Daten, in Allgemeines Statistisches Archiv, 3.1981, S. 273 - 291.

Arminger, G. (1982), Beiträge zu einem einheitlichen Modell sozialwissenschaftlicher Datenanalyse, Arbeitstitel eines unveröffentlichten Manuskripts, Wuppertal.

Bhapkar, V.P. (1966), A note on the equivalence of two criteria for hypotheses in categorical data, Journal of the American Statistical Association, Vol. 61, S. 228 - 235.

Birch, M.W. (1963), Maximum likelihood in three-way contingency tables, Journal of the Royal Statistical Society, Series B, Vol. 25, S. 220 - 233.

Bishop, Y.M.M., Fienberg St.E., Holland, P.W. (1975), Discrete multivariate analysis; theory and practice, Cambridge, Mass.

Bock, R.D. (1975), Multivariate statistical methods in behavioral research, New York.

Evers, M., Namboodiri, N.K. (1979), On the design matrix strategy in the analysis of categorical data, in K. Schuessler, Editor, Sociological Methodology 1979, San Francisco.

Fienberg, St., E. (1977), The analysis of cross classified categorical data, Cambridge, Mass.

GLIM Manual (1978), by Baker R.J. and Nelder J.A., Numerical Algorithm Group, Oxford.

Goethe, J.W. von (1796), Wilhelm Meisters Lehrjahre, Hrsg. Schmidt E: Vierter Band der Ausgabe des Insel Verlags, Leipzig 1914, S. 57.

Goodman, L.A. (1970), The multivariate analysis of qualitative data: interactions among multiple classifications, Journal of the American Statistical Society, Vol. 65, S. 226 - 256.

Goodman, L.A. (1972), A general model for the analysis of surveys, American Journal of Sociology Vol. 77, S. 1035 - 1086.

Goodman, L.A. (1978), Analyzing qualitative/categorical data: loglinear models and latent structure analysis, Cambridge, Mass.

Grizzle, J.E., Starmer, C.F. and Koch, G. (1969), Analysis of categorical data by linear models, Biometrics, Vol. 25, S. 489 - 504.

Haberman, S.J. (1974), The analysis of frequency data, Chicago.

Haberman, S.J. (1977), Log-linear models and frequency tables with small expected call counts, Annals of Statistics, Vol. 5, S. 1148 - 1169.

Holm, K. (1979), Das allgemeine lineare Modell, in Die Befragung 6, Hrsg. K. Holm, S. 1 - 261, München.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lee, T.C. (1980), The theory and practice of econometrics, New York.

Kendall, M., Stuart, A. (1979), The advanced theory of statistics,Vol. 4th Edition, London.

Küchler, M. (1979), Multivariate Analyseverfahren, Stuttgart.

Küchler, M., Hides, J.W., (1981), Economic Perceptions and the '76 and '80 Presidential Votes. Paper presented at ASA Meetings, Toronto.

Langeheine, R. (1980), Log-lineare Modelle zur multivariaten Analyse qualitativer Daten, eine Einführung.

Nelder, J.A. and Wedderburn, R.W.M. (1972), generalized linear models, Journal of the Royal Statistical Society, Series A, Vol. 135, S. 370 - 384.

Nelder, J.A. (1981), Lectures about general linear models, gehalten am Institut für höhere Studien, Wien.

NONMET Manual (1981), by Kritzer, H.M., University of Wisconsin, Madison, Wisconsin.

Rao, C.R. (1962), Efficient estimates and optimum inference procedures in large samples, Journal of the Royal Statistical Society, Series B, Vol. 24, S. 46 ff.

Scheffé, H. (1959), The analysis of Variance, New York.

Searle, S.R. (1971), Linear Models, New York.

$ACCURACY [Integer]

Anzahl der Stellen bei der Ausgabe
von Zahlen im G-Format
durch $CALCULATE $LOOK $PRINT
Voreinstelung: 4
zulaessig : nicht-negative Zahlen
groesser 9 => 9
0 bzw. blank bedeuten Voreinstellung

$ALIAS -

Umschalten auf Ausschluss oder Einschluss
von intrinsically aliased parameters
Voreinstellung: Ausschluss

$ARGUMENT macro arguments
--
Setzen von maximal 9 Argumenten fuer ein
definiertes Makro. Die Argumente brauchen erst
beim Aufruf des Makros definiert zu sein.
Es duerfen mehr Argumente definiert
als benutzt werden.
Die Argumente stehen in der Reihenfolge ihrer Nennung
fuer die formalen Argumente %1,%2 etc.
innerhalb des Makros. Zulassige Argumente sind:
Identifier
formales Argument des setzenden Makros (z.B. %1)
* = Zeichen fuer Beibehaltung einer Setzung
Fehlerhinweis: GLIM3 interpretiert bei manchen
Makro-Verschachtelungen falsch.

$CALCULATE Ausdruck
---
in der Form Operand [Operator Operand]en
Diese Direktive wertet arithmetische oder logische
Ausdruecke aus und weist die Ergebnisse Zieloperanden
zu. Dabei koennen Zieloperanden als Vektoren
implizit definiert werden. Der zuletzt ausgewertete
Ausdruck kann zusaetzlich protokolliert werden.

Operatoren sind + - * / ** und =

Operanden sind
1. Einfacher Operand : vorzeichenlose Zahl,
Skalar oder Vektor
2. Funktion, gefolgt von eingeklammerten Ausdruecken
(vgl. Anhang 1 GLIM-Funktionen)
3. Teilvektor, dargestellt durch einen Vektor,
gefolgt von einem eingeklammerten Index-
ausdruck, welcher entweder ein einzelner Index
oder ein Vektor von Indizes ist
4. Negierter Operand, bestehend aus dem
Minuszeichen, welchem ein Operand folgt
5. Klammerausdruck
6. Impliziter Operand, bestehend aus dem
Substitutionszeichen, welchem ein Makro-
name folgt. Der Makrostring tritt an

die Stelle des Operanden

Reihenfolge der Abarbeitung eines Ausdrucks:
1. Zuweisungen werden rechts beginnend nach links hin fortgesetzt
2. Innerhalb eines zu berechnenden und ggf. zuzuweisenden Ausdrucks gilt die Operatorenreihenfolge ** / * - + =

Jeder Ausdruck hat eine Laenge. Operanden eines Ausdrucks brauchen keine Laenge zu haben. Falls ein Operand eine Laenge hat, muessen alle uebrigen laengenbehafteten Operanden desselben Ausdrucks dieselbe Laenge haben, welche zur Laenge des Ausdrucks wird. Hat keiner der Operanden eines Ausdrucks eine Laenge, so hat auch der Ausdruck keine eigentliche Laenge.
Ein vorher definierter Ergebnisvektor muss dieselbe Laenge wie der ihm zuzuweisende Ausdruck haben. Ein vorher noch nicht definierter Ergebnisvektor erhaelt seine Laenge durch den ihm zuzuweisenden Ausdruck: diese Laenge ist entweder die definierte Laenge des Ausdrucks oder die Standardlaenge %NU, falls der Ausdruck keine definierte Laenge hat.
Die Operandenlaengen in einem Ausdruck bestimmen sich wie folgt.
1. Einfache Operanden: Zahlen, Skalare und Vektoren der Laenge 1 definieren keine Operandenlaenge. Ein Vektor der Laenge groesser 1 gibt seine Laenge dem Operanden
2. Die Laenge einer Funktion ist die Laenge ihres Argumentausdrucks
3. Die Laenge eines Teilvektors ist die Laenge seines Indexausdrucks.

Der Indexausdruck eines Teilvektors wird nach der ueblichen Auswertung gerundet. Zulaessige Indexwerte sind ganze Zahlen von 1 bis zur Laenge des indizierten Vektors sowie die Zahl 0. Ein Index 0 fuer den Ergebnisvektor bedeutet "keine Zuweisung", ein Index 0 auf der rechten Seite bedeutet die Zahl 0.

Falls der Gesamtausdruck keine Zuweisung enthaelt oder falls sein am weitesten links stehender Operator nicht die Zuweisung = ist, werden alle Werte des Gesamtausdrucks als Spaltenvektor gedruckt

```
$C          string
---         Kommentar, ohne das Direktivensymbol
            Fehler: Das Substitutionssymbol ? bzw. #
                    wird nicht als Kommentar angesehen!

$CYCLE      [Integer1 [Integer2]]
---         Integer1 = maximaler Iterationsschritt
                       (Voreinstellung: 10)
            Integer2 = jeder Integer2. Schritt und der
                       letzte Schritt werden gedruckt
                       (Voreinstellung: nur der letzte Schritt)
            ausserdem werden beim $FIT eines Standarmodells
            jeweils die %YV als Anfangswerte fuer die %FV
            gesetzt (vgl. jedoch $RECYCLE)

$DATA       [Integer] identifiers
---         definiert spaeter durch $DINPUT oder $READ
            zu lesende Vektoren der Laenge Integer. Falls
            die Laengenangabe fehlt, erhalten noch undefinierte
            Vektoren die Laenge des ersten Vektors und ggf.
            der noch nicht definierte erste Vektor die Standard-
            laenge.
            Fehler: ungleiche Laenge der Vektoren
            gueltig bis: Loeschung eines der genannten Vektoren
            Hinweis: max. 32 Vektoren in
            einer DATA-Anweisung

$DELETE     identifiers
---         loescht Identifier
            Es duerfen noch nicht definierte Identifier
            genannt werden. Waehrend eines FIT duerfen
            keine Modellmakros oder Modellvektoren geloescht
            werden. Auch Makros aus der lfd. Programmhierarchie
            duerfen nicht geloescht werden.

$DINPUT     Integer1 [Integer2]
----        liest Zahlen von Kanal Integer1 mit der maximalen
            Satzlaenge Integer2 (32 bis 299)
            gemaess der Definition durch $DATA.
            Das Lesen erfolgt wie durch $READ, aber
            statt vom lfd. Eingabekanal jetzt von Kanal
            Integer1, welcher noch nicht in der Programm-
            hierarchie sein darf.
            Voreinstellung fuer Integer2:
            Satzlaenge des Primaereingabekanals (vgl. $ENV C)
```

```
$DISPLAY      letters
--            zeigt Ergebnisse des letzten FITs mit noch
              verfuegbaren Ergebnissen.
              letter:
              A  wie E, incl. intrinsically aliased Parameter
              C  Korrelationen der Parameterschaetzwerte
              D  Anpassungsmass (Deviance) und Freiheitsgrade
              E  Parameterschaetzwerte, ihre Standardfehler
                 und Definitionen, incl. extrinsically aliased
                 Parameters
              L  Formel fuer linearen Praediktor
              M  alle Modell-Spezifikationen
              R  die Y-Variate, ihre Schaetzwerte, generali-
                 sierte Residuen und ggf. binomiale Nenner
              S  Standardfehler der Differenzen geschaetzter
                 Parameter
              T  generalisierte Inverse der SSP-Matrix
              U  wie E, excl. extrinsically aliased Parameters
              V  Covarianz-Matrix der Parameterschaetzwerte
              W  wie R, aber abhaengig von der Maske %RE

$DUMP         [Integer]
---           Binaerdump auf Kanal Integer ab lfd. Position
              Voreinstellung: Standarddumpkanal vgl. $ENV C

$ECHO         -
---           Umschalten auf Protokollierung
              bzw. Nichtprotokollierung der Eingabe
              Voreinstellung: interaktiv = Nichtprotokollierung
                              batch      = Protokollierung
              Makros und eingelesene Daten werden nie protokolliert

$EDIT         [Integer1 [Integer2]] Vektoren Zahlen
---           weist Vektoren Zahlen fuer die Elemente
              Integer1 bis Integer2 zu
              Voreinstellung: Integer1 = 1
                              Integer2 = Vektorlaenge
              Fehler: die erste Zahl darf weder mit dem
              Minuszeichen noch mit dem Punkt beginnen

$ENDMAC       -
--            1. Beendet ein Macro, wenn innerhalb eines Makros
              2. Beendet einen Job, wenn ausserhalb eines Makros
```

```
$ENVIRONMENT [Integer] letters
----         Protokollierung des Programmzustandes
             auf Kanal Integer oder den lfd. Ausgabekanal
             letter:
             C  channel
                 Kanaele fuer Eingabe, Ausgabe und Dump
             D  directory
                 benutzerdefinierte Identifier und benutzte
                 Systemvektoren sowie deren Speicherplaetze
             I  implementation
                 implementationsabhaengige Eigenschaften
             P  program-control stack
                 Programmebenen
             R  pseudo-random-number
                 Startwerte fuer den Standardzufallsgenerator
             S  system
                 vom GLIM-System benutzter Speicherplatz
             U  usage
                 Speicherplatz der Daten, Identifier, Vektoren
                 Modellterme, Programmebene

$ERROR       Name [Identifier]
---          systemdefinierte Dichtefunktionen
             Name:
             Binomial binomialer Nenner
             -
             Gamma
             -
             Normal
             -
             Poisson
             -
             ein zusaetzlicher Identifier ist bei Binomial
             und nur bei Binomial anzugeben

$EXIT        [Integer]
---          Reduziert die Programmebene um Integer
             Ebenen unbedingt (vgl. $SKIP)

$EXTRACT     Identifiers
----         weist Werte aus der SSP-Matrix den
             Identifiers %VC, %PE oder %VL zu

$FACTOR      [Integer1] [Identifier Integer]s
---          definiert einen Vektor der Laenge
             Integer1 als Factor (Nominale Groesse)
             mit der maximalen Auspraegung Integer.
             Eine fehlende Vektorlaenge Integer1 fuer einen
             noch nicht definierten Vektor wird durch die
             Standardlaenge ersetzt. Eine angegebene Vektor-
             laenge Integer1 fuer einen schon definierten
             Vektor muss der bekannten Vektorlaenge gleichen
             (Verbot der Redefinition von Vektorlaengen!).
             Die maximale Auspraegung Integer darf rede-
             finiert werden. Mit $FACTOR und $VARIATE koennen
             also metrische in kategoriale Groessen ver-
             wandelt werden und umgekehrt.
```

$FINISH -
---- bedeutet Dateiende beim Aufsuchen von Subfiles in einer Sekundaerdatei ($INPUT). Wenn das gesuchte Subfile noch nicht gefunden wurde, bewirkt das erste Erreichen von $FINISH ein Rewind der Datei und die Fortsetzung des Suchvorgangs laengstens bis zum zweiten Auffinden von $FINISH. $FINISH darf nicht auf dem Primaereingabekanal gefunden werden.

$FIT [Modell-Formel]
-- berechnet die Anpassung an das durch $ERROR und $LINK bzw. $OWN, $YVARIATE, $WEIGHT, $OFFSET und $SCALE definierte Modell. Die Berechnung kann weiterhin beeinflusst werden durch $CYCLE, $RECYCLE und $ALIAS.
Eine Modellformel kann bestehen aus
Operatoren . / * , + - -/ */
Operanden Vektoren
und Klammern ( )
durch $FIT koennen Modellformeln entweder gesetzt oder veraendert werden.
Voreinstellung: %GM
Durch Loeschen eines in der Modellformel enthaltenen Vektors wird die Voreinstellung wiederhergestellt.

$FORMAT
-- definiert auf der folgenden Zeile das Format zu lesender Daten
leer oder FREE bedeutet freies Format, d.h. Trennung durch Blank(s) oder Zeilenwechsel
(Fortran-Format)
Voreinstellung: FREE
Fehler: innerhalb eines Makos darf mit $READ nur in freiem Format gelesen werden.
Hinweis: stimmen die einzulesenden Daten nicht mit dem Format ueberein, so wird mit einer Fortran-Fehlermeldung GLIM beendet.
Nach der Klammer zu des Fortran-Formats sollten noch 4 Blanks eingegeben werden

$HELP -
-- schaltet auf kurze oder ausfuehrliche Fehlermeldung
Voreinstellung: ausfuehrlich

```
$INPUT      Integer1 [Integer2] [Subfiles]
--          liest Direktiven von Kanal Integer1 (1 bis 99)
            mit der maximalen Satzlaenge Integer2 (30 bis 299)
            entweder ab dem naechsten Satz oder ab
            $SUBFILE Subfile. Wenn vor dem gesuchten
            Subfile ein $FINISH gefunden wurde, wird die
            Suche vom Dateianfang maximal bis $FINISH
            fortgesetzt.
            Das Ende des Subfiles mit Rueckkehr zum
            aufrufenden Kanal geschieht durch $RETURN, $FINISH,
            $EXIT 1 oder $SKIP 1. Weiter zurueckreichende
            Beendigungen erfolgen durch $EXIT >1 oder
            $SKIP >1 oder $END.
            Fehler: Kanal Integer1 in lfd. Programmhierarchie
            Hinweis: fehlt ein $FINISH, so koennen
            Endlosschleifen entstehen

$LINK       letter [Zahl]
---         definiert die Link-Funktion fuer eine durch
            $ERROR definierte Dichtefunktion
            letter:
            C  complementary log-log
            E  number  exponent
            G  logit
            I  identity
            L  log
            P  probit
            R  reciprocal
            S  square root
            Voreinstellung: Der natuerliche Parameter
            der Dichtefunktion
            Fehler: nicht jede Zuordnung von $ERROR
            und $LINK erlaubt.

$LOOK       [Integer1 [Integer2]] Vektoren oder
$LOOK       Skalare
--          druckt spaltenweise entweder Teile der
            angegebenen Vektoren mit den Indexgrenzen
            Integer1 bis Integer2 oder die Skalare.
            Beim Druck von Teilvektoren wird eine Spalte
            mit den Zeilenindizes vorangestellt.
            Falls die definierte Zeilenlaenge nicht ausreicht,
            wird rechts abgeschnitten.

$LSEED      [Integer1 [Integer2 [Integer3]]]
---         Startwerte fuer lokalen Zufallsgenerator (wie $SSEED)

$MACRO      identifier string
--          definiert ein Macro mit Namen identifier
            und Inhalt String oder ueberschreibt
            das Makro mit neuem Inhalt
            Fehler: Darf nicht innerhalb eines Makros
            verwendet werden.
```

$OFFSET [identifier]
--
identifier bezeichnet denjenigen Vektor, der beim $FIT zum linearen Praediktor hinzuaddiert wird.
Voreinstellung: undefiniert

$OUTPUT [Integer1 [Integer2 [Integer3]]]
---
definiert den Ausgabekanal Integer1 (1 bis 99) mit der max. Satzlaenge Integer2 (52 bis 132) und der Anzahl Saetze Integer3 (>5 bei $PLOT) fuer $DISPLAY $LOOK $PLOT $PRINT sowie fuer alle Warnungen und alle automatischen modellbezogenen Meldungen. Fehlermeldungen werden dagegen immer auf den Standardausgabekanal gegeben.

$OWN Identifier1 Identifier2 Identifier3 Identifier4
---
definiert ein benutzereigenes Modell durch 4 Makros. Die Verbindungsfunktion wird definiert durch Identifier1 und Identifier2, die Dichtefunktion unabhaengig von der Verbindungsfunktion durch Identifier3 und Identifier4.
Das Modell bleibt bestehen bis $ERROR oder $OWN
  Identifier1 berechnet %FV aus %LP
  Identifier2 berechnet %DR
  Identifier3 berechnet %VA
  Identifier4 berechnet %DI

$PAUSE -
---
unterbricht GLIM, um Betriebssystem-Kommandos zwischendurch ausfuehren zu koennen
(nicht implementiert)

$PLOT Vektoren Vektor1
--
zeichnet ein Scattergramm mit max. 9 Vektoren gegen Vektor1.
Falls %RE Werte hat, werden diejenigen Vektorkomponenten unterdrueckt, deren %RE-Werte 0 sind.
Das Anfangszeichen eines Vektors kennzeichnet jeden Punkt. Fallen mehrere Punkte zusammen, werden die Punkte durch ihre Haeufigkeit (max. 9) bezeichnet.

$PRINT [Items]
---
druckt Zahlen und Texte, Zeilenwechsel und Seitenwechsel auf dem lfd. Ausgabekanal.
Falls die lfd. Zeile ueberfuellt wird, wird die Ausgabe auf der naechsten Zeile fortgesetzt.
Item:
 : neue Zeile
 / neue Seite
 *integer temporaere Stellenanzahl anstatt $ACCURACY
 'string' Text ohne das Direktivensymbol
 identifier entweder Zahl (Skalar), Zahlen (Vektor) oder Text (Makroinhalt); der identifier darf ein formales Argument sein bzw. ueber einen Makroinhalt substituiert werden.

$READ Zahlen
liest Zahlen vom lfd. Eingabekanal mit der definierten max. Satzlaenge. Die gelesenen Zahlen werden zeilenweise abgelegt in die durch $DATA aus Spaltenvektoren definierte Matrix. Die gelesenen Zeilen duerfen kuerzer oder laenger als die Matrixzeilen sein. Das zeilenweise Lesen endet, wenn die Matrix gefuellt ist.
Formatiertes Lesen mit $READ ist nur ausserhalb von Makros gestattet.

$RECYCLE [Integer1 [Integer2]]
wie $CYCLE, jedoch werden beim FIT eines Standard-Modells keine neuen Anfangswerte fuer die %FV gesetzt

$REINPUT Integer1 [Integer2] [Subfiles]
wie $INPUT,
jedoch vorher automatisches Rewind.

$RESTORE [Integer]
liest einen frueheren DUMP von Kanal Integer
Voreinstellung: Standarddumpkanal (vgl. $ENV C)
Hinweis: versucht man ueber das Datenende hinaus zu lesen, so wird mit einer Fortran-Fehlermeldung GLIM beendet.

$RETURN -
beendet das durch $INPUT oder $REINPUT veranlasste Lesen vom Eingabekanal bzw. das durch $SUSPEND veranlasste temporaere Lesen vom Standardeingabekanal

$REWIND [Integer]
setzt den Kanal Integer auf den Anfang

$SCALE [Zahl]
setzt Zahl als Anfangswert fuer Scaleparameter:
0 oder leer bedeutet iteratives Ersetzen durch den jeweiligen Quotienten aus Deviance und Freiheitsgrad.
Positive Zahl bedeutet Fixierung auf den Wert Zahl.
Voreinstellung: durch $ERROR entweder auf
0 (Normal-,Gammaverteilung)
oder 1 (Binomial-,Poissonverteilung)

$SKIP [Integer]
wie $EXIT
mit dem Unterschied, dass die letzte zu verlassende Programmebene nur bedingt verlassen wird, falls sie ein durch $WHILE aufgerufenes Makro ist.

```
$SORT   Vektor1 [Vektor2 oder Integer2 [Vektor3 oder Integer3]]
--          allgemeine Sortierroutine.
            Das Ergebnis Vektor1 geht aus Vektor2 dadurch hervor,
            dass auf Vektor2 jene Permutation angewandt wird,
            die zur aufsteigenden Sortierung von Vektor3
            benoetigt wuerde:
                                    -1
             Vektor1 = Vektor2(inverser Rang   (Vektor3))
             Vektor2 = Vektor1(Rang(Vektor3))
            Spezielle Sortierroutinen sind:
             $SORT Vektor1 1 Vektor3
             erzeugt  Vektor1 = inverser Rang(Vektor3)
             $SORT Vektor1 Vektor2 k
             erzeugt  Vektor1 = um k-1 Positionen circular
                                verschobener Vektor2 (lag)
             $SORT Vektor1 Vektor2 [Vektor2]
             erzeugt  Vektor1 = aufsteigend sortierter Vektor2
             Die Raenge von Vektor3 in Vektor1 abzuspeichern,
             erfordert zwei Sortieraufrufe:
             $SORT Vektor1 1 Vektor3 : Vektor1 1 Vektor1

$SSEED      [Integer1 [Integer2 [Integer3]]]
---         setzt die Startwerte des Standard-Zufallsgenerator

$STOP       -
---         beendet GLIM

$SUBFILE    Identifier
---         definiert am Anfang einer Zeile den Anfang eines
            Subfiles mit Namen Identifier, ein Subfile
            endet mit $RETURN oder $FINISH
            Fehler: falls Identifier bekannt und kein Subfile

$SUSPEND    -
----        ruft den Standardeingabekanal auf,
            von welchem mit $RETURN, $EXIT oder $SKIP
            zurueckgekehrt wird.
            $END bedeutet hier end of job !
            $FINISH auf dem Standardeingabekanal ist verboten.

$SWITCH     scalar macros
---         bedingter Aufruf eines Macros:
            scalar bezeichnet nach Rundung die Reihenfolgenummer
            des aufzurufenden Makros; falls scalar < 1 oder
            > Anzahl Makros, erfolgt kein Aufruf

$UNITS      Integer
---         Integer ist Standardlaenge der Vektoren
            und Stichprobenumfang fuer den FIT

$USE        macro
--          unbedingtes Aufrufen des Makros macro
```

$VARIATE [Integer] Idendifiers
-- definiert Vektoren der Laenge Integer
Eine fehlende Vektorlaenge Integer fuer einen noch nicht definierten Vektor wird durch die Standardlaenge ersetzt. Eine angegebene Vektorlaenge Integer fuer einen schon definierten Vektor muss der bekannten Vektorlaenge gleichen (Verbot der Redefinition von Vektorlaengen!).

$WARNING
--- ermoeglicht oder unterdrueckt Warnungen
Voreinstellung: moeglich

$WEIGHT [Identifier]
-- Vektor Identifier enthaelt nicht-negative a-priori-Gewichte fuer FITs. Es erfolgt keine automatische Normierung der Gewichte.

$WHILE skalar macro
--- ruft das Makro macro solange auf, bis scalar genau 0 (!) ist.
Falls scalar mit 0 beginnt oder falls der Inhalt des Makros leer ist, erfolgt kein Aufruf.

$YVARIATE Identifier
-- der Vektor Identifier wird als abhaengige Variable definiert

ANHANG 1 G L I M - FUNKTIONEN

```
%ANG(X))     = arcsin(sqrt(X))  ; 0 .le. X .le. 1

%EXP(X)     = e**X

%LOG(X)     = log(X) zur Basis e

%SIN(X)

%SQRT(X)

%NP(X)       = Integral der Standard-Normalverteilung
                  von -Unendlich bis X

%ND(X)       = Umkehrfunktion von %NP ; 0 .le. X .le. 1

%TR(X)       = ganzzahliger Anteil von X
               durch Abrunden des Betrages von X

%GL(k,n)     = fuellt nacheinander einen Vektor mit Bloecken
               aus je n natuerlichen Zahlen, beginnend bei 1 und
               spaetestens endend bei k. Falls eine vordefinierte
               Laenge des Vektors groesser ist als k*n,
               wird die Zuweisung zyklisch fortgesetzt.
               Die Argumente k bzw. n duerfen auch Vektoren sein.

%CU(X)       = kumulierte Summen

%LT(A,B)     = 1, falls A kleiner als B      ; = 0, falls nicht

%LE(A,B)     = 1, falls A kleiner/gleich B   ; = 0, falls nicht

%EQ(A,B)     = 1, falls A gleich B           ; = 0, falls nicht

%NE(A,B)     = 1, falls A ungleich B         ; = 0, falls nicht

%GE(A,B)     = 1, falls A groesser/gleich B  ; = 0, falls nicht

%GT(A,B)     = 1, falls A groesser als B     ; = 0, falls nicht

%IF(bed,wt,wf)
            bed = Bedingung (1 fuer wahr, 0 fuer falsch)
             wt = Wert falls Bedingung wahr
             wf = Wert falls Bedingung falsch

%SR(n)          standard random    und
%LR(n)          local random       generieren Zufallszahlen
      n .lt. .5  gleichverteilte Zahlen zwischen 0 und 1
      n .ge. .5  gleichverteilte Zahlen auf 0,1,2, ... ,[n + .5]
```

ANHANG 2 G L I M - SYSTEM-VEKTOREN

%FV fitted values by $FIT, deleted by $UNITS

%LP linear predictors

%WT iterative weights

%WV working vector = %LP + (%YV - %FV) * %DR
i.e the adjusted linear predictor

%YV dependent variate

%BD binomial denominator

%PW prior weights

%OS declared offset

%DR derivative d%LP/d%FV by user defined macro2

%VA variance function, i.e. second derivation of the
B-function by its natural parameter (not the variance!)

%DI deviance increments

%GM values 1 for dummy-variable grand mean

%PE parameter estimates of length %PL due to
not intrinsically aliased parameters
to be extracted

%VC variance-covariance matrix of length %ML
due to %PE
to be extracted

%VL variances of the linear predictors,
to be extracted

%RE zero elements suppress output of related units
in $PLOT and $DISPLAY W

ANHANG 3 G L I M - SYSTEM-SKALARE

%JN job number

%NU number of units

%DV deviance

%DF degrees of freedom

%X2 generalised pearson chisquare statistic

%SC scale parameter, either fixed or
replaced by mean deviance

%CL current level of the programm-control stack

%ML number of elements in the covariance matrix %VC

%PL number of parameter estimates %PE

%PI 3.14159

```
$SUBFILE DEBS
$C
$C
$C
$C        G.ENERALISED                              D.EMONSTRATION OF
$C        L.INEAR                                   E.STIMATION RESULTS FROM
$C        I.NTERACTIVE                              B.INOMIALLY DISTRIBUTED
$C        M.ODELLING                                S.IMULATED SAMPLES
$C
$C
$C
$C    ***************        1. VERSION RESTRICTED TO
$C    *             *
$C    *   D E B S   *        COEFFICIENTS OF ESTIMATED REGRESSAND Y
$C    *             *        FORMULATED EITHER AS Y=A+B*X OR AS Y=(X-M)/S
$C    *      1      *        AND THEIR PAIRWISE CORRELATION COEFFICIENTS
$C    *             *
$C    ***************
$C
$C    BY H. BUSSE          STARTED  AT MANNHEIM  19.NOV.1981
$C                         FINISHED IN BERLIN    11.JAN.1982
$C
$C
$C
$C    AIM OF THIS PROGRAM
$C    -------------------
$C
$C    THE FOLLOWING GLIM-PROGRAM SHALL ENABLE THE USER TO GET FREQUENCY
$C    DISTRIBUTIONS OF SOME ESTIMATED FUNCTIONS OCCURRING IN SIMPLE LINEAR
$C    REGRESSION ANALYSIS OF BINOMIALLY DISTRIBUTED DATA.
$C
$C    THE DATA ARE TO BE GENERATED BY MONTE-CARLO-METHOD DEPENDING ON GIVEN
$C    DISTRIBUTION PARAMETERS AND CONTROL PARAMETERS.
$C
$C    THE PROGRAM CONSISTS OF SOME MACROS SUITABLE EITHER TO EXECUTE
$C    THE WORK TOTALLY OR TO INSPECT RESULTS ADDITIONALLY.
$C
$C    GENERALLY, THE PROGRAM USE MAY CONSIST OF PARAMETRIZATION AND
$C    FOLLOWING BATCH-LIKE RUN OF THE PROGRAM USING DEFAULT VALUES
$C    FOR ALL CONTROL PARAMETERS.
```

```
$C
$C
$C    CONTROL PARAMETERS
$C    ------------------
$C
$C    1.  NUMBER %NU OF SUBGROUPS IN EACH SAMPLE AND
$C        NUMBER %G  OF SAMPLES TO BE TREATED
$C        TASK NUMBER %Q
$UNITS 5
$CALCU %G=100
     : %Q=1
$C
$C    2.  VECTOR NN OF SUBSAMPLE NUMBERS AND
$C        VECTOR PP OF SUBSAMPLE PROBABILITIES
$DELET NN PP
$CALCU NN=PP=0
$EDIT  NN   10  10  10  10  10
$EDIT  PP 0.15 .35 .55 .75 .95
$C
$C    3.  PARAMETERS OF THE UNDERLYING DISTRIBUTION FUNCTION DEFINING PP:
$C        %I   TYPE OF LINK FUNCTION Y(P) WHILE GENERATING SAMPLES
$C        %J   TYPE OF LINK FUNCTION Y(P) WHILE ANALYZING  SAMPLES
$C              1   PROBIT                                MEDIAN(Y)= 0
$C              2   LOGIT                                 MEDIAN(Y)= 0
$C              3   COMPLENTARY LOGLOG = 'CLOGLOG'        MEDIAN(Y)= LOG(LOG(2))
$C        %M   MEDIAN
$C        %S   STANDARD DEVIATION
$CALCU %I=1 : %J=1
     : %M=2 : %S=1
$C
$C    4.  STARTING VALUES CSR FOR RANDOM GENERATOR %SR
$VARIA 3 CSR
$CALCU   CSR=0
$EDIT    CSR 8 12 31
$C
$C    5.  OUTPUT CONTROL EACH CONCERNING 3 TYPES OF PAIRED RESULTS:
$C         1  REGRESSION COEFFICIENTS A AND B OF Y=A+B*X
$C         2  MEAN M AND STANDARD DEVIATION S OF Y=(X-M)/S
$C         3  CORRELATION COEFFICIENTS BETWEEN A AND B
$C                                  AND BETWEEN M AND S
$C         CCB = BIVARIATE GRAPHICS
$C         CCF = NUMBER OF FREQUENCY CLASSES
$C         CCH = HISTOGRAMS CONTROLLED BY THE SUM OF POSSIBLE ITEMS
$C                 1 FIRST  SINGLE
$C                 2 SECOND SINGLE
$C                 4 BOTH   TOGETHER
$C         CCN = NUMERICAL PRESENTATION
$VARIA 3 CCB CCF CCH CCN CCS
$CALCU   CCB=CCF=CCH=CCN=0
$EDIT    CCB  1  1  1
    :    CCF  0  0 15
    :    CCH  0  3  4
$C  :    CCN  1  1  0
$C          AB MS RR
$C
$C    6.  SET OUTPUT CHANNELS AND OUTPUT FORMAT
$C         1  CHANNEL FOR FORCED OUTPUT
$C         2  CHANNEL FOR WISHED OUTPUT
$C         3  MAXIMAL WIDTH  (BETWEEN 52 AND 132)
$C         4  MAXIMAL HEIGHT (AT LEAST 6        )
$C         5  SCRATCH VALUE
$VARIA 5  VOU
$CALCU    VOU=0
$EDIT 1 4 VOU 20  6  117  41
```

```
$C
$C
$C   TEXTS CONCERNING SIMULATED FUNCTIONS FOR GENERALIZED OUTPUT
$C
$MACRO BLAT                                          $ENDMA
$C
$MACRO ABAT A-COEFFICIENT OF Y=A+B*X                 $ENDMA
$MACRO ABBT B-COEFFICIENT OF Y=A+B*X                 $ENDMA
$C
$MACRO MSMT M-COEFFICIENT OF Y=(X-M)/S               $ENDMA
$MACRO MSST S-COEFFICIENT OF Y=(X-M)/S               $ENDMA
$C
$MACRO ABRT CORRELATION(A,B) IN Y=A+B*X              $ENDMA
$MACRO MSRT CORRELATION(M,S) IN Y=(X-M)/S            $ENDMA
$C
$C
$C   LETTERS OF ORDINATES TO BE PLOTTED
$MACRO L1TX L $ENDMA
$MACRO L2TX S $ENDMA
$MACRO RTEX R $ENDMA
$C
$C
$C   DEFINITIONS OF THE LINK FUNCTION
$MACRO LIK1 $LINK P $ARGUM HEAD * LIT1 $ENDMA
$MACRO LIK2 $LINK G $ARGUM HEAD * LIT2 $ENDMA
$MACRO LIK3 $LINK C $ARGUM HEAD * LIT3 $ENDMA
$C
$C   LINK-DEPENDENT TEXTS
$MACRO LIT1 PROBIT   $ENDMA
$MACRO LIT2 LOGIT    $ENDMA
$MACRO LIT3 CLOGLOG  $ENDMA
$C
$C   LINK-DEPENDENT CALCULATIONS OF PREDICTORS
$MACRO LIY1 $CALCU YY=%ND(PP)             $ARGUM HEAD LIT1 $ENDMA
$MACRO LIY2 $CALCU YY=%LOG(PP/(1-PP))     $ARGUM HEAD LIT2 $ENDMA
$MACRO LIY3 $CALCU YY=%LOG(-%LOG(1-PP))   $ARGUM HEAD LIT3 $ENDMA
$C
$C   LINK-DEPENDENT CALCULATIONS OF PROBABILITIES
$MACRO LIP1 $CALCU PP=%NP(YY)             $ENDMA
$MACRO LIP2 $CALCU PP=1/(1+%EXP(YY))      $ENDMA
$MACRO LIP3 $CALCU PP=1-%EXP(%EXP(YY))    $ENDMA
$C
$C
$C   LINK-DEPENDENT PREDICTOR VALUES OF THE MEDIAN
$VARIA 3 MEDY
$CALCU   MEDY=0
     :   MEDY(3)=%LOG(%LOG(2))
$C
$C ..........................................................
$C
$MACRO TOTA
$C      T O T A    SOLUTION OF THE WHOLE PROBLEM
$C   REDEFINE VARIABLE FUNCTIONS
$USE   REDF
$C   GENERATE AND ANALYSE 20 SAMPLES
$USE   GENA
$OUTPU $
$C   OUTPUT
$USE   RESP
$PRINT /
$OUTPU $
$ENDMA
```

```
  H.BUSSE     11.1.82          BEISPIEL EINES GLIM-PROGRAMMS
$C
$C
$MACRO REDF
$C      R E D F    REDEFINE VECTORS OF VARIABLE DIMENSION AND
$C                 COMPUTE OTHER FUNCTIONS OF CONTROL PARAMETERS
$C
$C   SUBSEQUENT FORCED OUTPUT INTO AUXILIARY CHANNEL
$CALCU %X=VOU(1)
$OUTPU %X
$C
$C   DELETE AND REDEFINE SAMPLE VECTORS AND GENERAL SCRATCH VECTORS
$DELET  YY XX RR ZWI1 ZWI2
$VARIA  YY XX RR
    : 13 ZWI1
    :  7 ZWI2
$C
$C   TRANSFORM (%M,%S) INTO (%A,%B)
$CALCU %B=1/%S
      : %A=MEDY(%I)-%B*%M
$C
$C   COMPUTE EQUIDISTANT REGRESSORS XX DEPENDING ON PROBABILITIES PP
$C   AND ON TYPE %I OF GENERATED LINEAR PREDICTOR
$SWITC %I LIY1 LIY2 LIY3
$CALCU XX=%M+%S*(YY-MEDY(%I))
$CALCU %X=%TR((%NU+1)/2)
      : %Y=XX(%X) : %Z=XX(%NU)
      : ZWI1(3)=%NU-%X
      : ZWI1(2)=%Z-%Y
      : ZWI1(1)=%Y*%NU-%Z*%X
$CALCU XX=(ZWI1(1)+ZWI1(2)*%GL(%NU,1))/ZWI1(3)
$C
$C   RECOMPUTE PREDICTORS YY AND PROBABILITIES PP
$CALCU YY=%A+%B*XX
$SWITC %I LIP1 LIP2 LIP3
$C
$C   MODEL DEFAULTS
$YVARI RR
$ERROR BI NN
$SWITC %J LIK1 LIK2 LIK3
$WEIGH
$OFFSE
$C
$C   INITIALIZE RANDOM NUMBER GENERATOR %SR
$CALCU %X=CSR(1) : %Y=CSR(2) : %Z=CSR(3)
$SSE    %X %Y %Z
$C
$C   COMBINE ALL OUTPUT REQUESTS
$CALCU CCS=CCB+%NE(CCE,0)*(CCH+CCM)
$C
$C   REDEFINE SIMULATION RESULTS
$DELET     ABAE ABBE ABRE
           MSME MSSE MSRE
$VARIA %G  ABAE ABBE ABRE
           MSME MSSE MSRE
$C
$C   REDEFINE VECTORS FOR TRANSFORMED RESULTS
$DELET     GM1 GM2 GF1 GF2
$VARIA %G  GM1 GM2 GF1 GF2
$ENDMA
```

```
$C
$C
$MACRO GENA
$C      G E N A   GENERATE AND ANALYZE %G SAMPLES AND ADJUST RESULTS
$CALCU %K=%G
$WHILE %K MOGA
$DELET PR
$ENDMA
$C
$C
$MACRO RESP
$C      R E S P   RESULT PRINTS
$C
$C   SUBSEQUENT WISHED OUTPUT INTO MAIN CHANNEL
$CALCU %Y=VOU(3) : %Z=VOU(4)
$USE    SOCH
$C
$C   OUTPUT CONCERNING PARAMETERS A AND B IN PREDICTOR FORMULA Y=A+B*X
$CALCU %O=1
$ARGUM PRES ABAE ABBE ABAT ABBT L1TX
$USE    PRES
$C
$C   OUTPUT CONCERNING PARAMETERS M AND S IN PREDICTOR FORMULA Y=(X-M)/S
$CALCU %O=2
$ARGUM PRES MSME MSSE MSMT MSST L2TX
$USE    PRES
$C
$C   OUTPUT CONCERNING CORRELATION COEFFICIENTS COR(A,B) AND COR(M,S)
$CALCU %O=3
$ARGUM PRES ABRE MSRE ABRT MSRT RTEX
$USE    PRES
$ENDMA
$C
$C
$MACRO MOGA
$C      M O G A   MONTE-CARLO GENERATION AND ANALYSIS OF ONE SAMPLE
$C       %K  SAMPLE INDEX
$C
$C   GENERATE A VECTOR RR OF %NU NUMBERS OF REAGENTS DUE TO
$C   PROBABILITIES PP AND SUBSAMPLE NUMBERS NN
$CALCU %L=%NU
$WHILE %L GENS
$C
$C   FIT AND STORE ESTIMATED PARAMETERS
$FIT    XX
$EXTRA %PE %VC
$CALCU ABAE(%K)=%PE(1)
      : ABBE(%K)=%PE(2)
      : MSME(%K)=%X=MEDY(%J)-%PE(1)/%PE(2)
      : MSSE(%K)=1/%PE(2)
$C
$C   CORRELATION BETWEEN A AND B
$CALCU ABRE(%K)=%VC(2)/%SQRT(%VC(1)*%VC(3))
$C   CORRELATION BETWEEN M AND S
$CALCU MSRE(%K)=(%VC(2)+%X*%VC(3))
                /%SQRT(%VC(3)*
                (%VC(1)+2*%X*%VC(2)+%X*%X*%VC(3)))
$C   COUNT SAMPLE INDEX TOWARDS ZERO
$CALCU %K=(%K-1)*%GT(%K,1)
$ENDMA
```

```
 H.BUSSE     11.1.82          BEISPIEL EINES GLIM-PROGRAMMS
$C
$C
$MACRO GENS
$C      G E N S    GENERATE THE NUMBER RR OF REAGENTS IN THE
$C                 CURRENT SUBSAMPLE OF INDEX %L
$C
$C  GENERATE A VECTOR ZZ OF UNIFORMLY DISTRIBUTED CONTINUOUS
$C  RANDOM NUMBERS FROM (0,1)
$DELET     ZZ
$CALCU %X=NN(%L)
$VARIA %X ZZ
$CALCU     ZZ=%SR(0)
$C
$C  COUNT THOSE RANDOM NUMBERS ZZ NOT EXCEEDING PROBABILITY PP
$CALCU RR(%L)=%CU(%LE(ZZ,PP(%L)))
$DELET ZZ
$C
$C  REGARD IMPLEMENTATION ERROR ON TR440
$C  WHEN COUNTING THE SUBSAMPLE INDEX TOWARDS ZERO
$CALCU %L=(%L-1)*%GT(%L,1)
$ENDMA
$C
$C ------------------------------------------------------------
$C
$MACRO PRES
$C      P R E S    MAIN MACRO FOR OUTPUT CONTROL
$C       %1  FIRST  VECTOR OF ESTIMATED PARAMETERS OR THEIR FUNCTIONS
$C       %2  SECOND VECTOR OF ESTIMATED PARAMETERS OR THEIR FUNCTIONS
$C       %3  TEXT CONCERNING %1
$C       %4  TEXT CONCERNING %2
$C       %5  LETTER OF ORDINATE VECTOR TO BE PLOTTED
$C       %0  MAIN CONTROL CLASS
$C            1  A AND B
$C            2  M AND S
$C            3  CORRELATIONS
$C
$C  ANY OUTPUT REQUESTED ?
$CALCU %X=%EQ(CCS(%0),0)
$EXIT  %X
$C
$C  BIVARIATE GRAPHIC REQUESTED ?
$CALCU %X=%NE(CCB(%0),0)
$ARGUM    GRAP %1 %2 %3 %4  %5
$SWITC %X GRAP
$C
$C  FURTHER OUTPUT REQUESTED ?
$CALCU %X=%EQ(CCF(%0)*(CCH(%0)+CCN(%J)),0)
$EXIT  %X
$C
$C  SET EXTREME VALUES
$ARGUM MIMA %1 ZWI1
$USE   MIMA
$ARGUM MIMA %2 ZWI2
$USE   MIMA
$CALCU ZWI1(4)=ZWI2(1)
     : ZWI1(5)=ZWI2(2)
     : ZWI1(6)=ZWI2(3)
```

```
$C
$C
$C   CONDITIONAL NUMBER %H OF FREQUENCY CLASSES
$CALCU %H=CCF(%O)
      : %H=%IF(%GT(%H,3),%H,3)
      : %X=VOU(3)-25
      : %H=%IF(%LT(%H,%X),%H,%X)
$C
$C   CLASSIFY BOTH VECTORS %1 AND %2
$DELET    KM1 KF1 KM2 KF2
$VARIA %H KM1 KF1 KM2 KF2
$ARGUM CLAF %1 KM1 KF1 GM1 GF1 %G %H
$USE   CLAF
$ARGUM CLAF %2 KM2 KF2 GM2 GF2
$USE   CLAF
$C
$C   HISTOGRAM DUE TO FIRST GIVEN VECTOR %1
$CALCU %X=CCH(%O)-2*%TR(CCH(%O)/2)
$ARGUM    HIS1 GF1 GM1 %3 ZWI1 %5

$SWITC %X HIS1
$C
$C   HISTOGRAM DUE TO SECOND GIVEN VECTOR %2
$CALCU %X=%TR(CCH(%O)/2)
      : %X=%X-2*%TR(%X/2)
$ARGUM    HIS1 GF2 GM2 %4 ZWI2
$SWITC %X HIS1
$C
$C   COMMON HISTOGRAM OF BOTH GIVEN VECTORS %1 AND %2
$CALCU %X=%NE(%TR(CCH(%O)/4),0)
$ARGUM    HIS2  %1 %2 %3 %4 ZWI1 %5
$SWITC %X HIS2
$C
$C   NUMERICAL PRESENTATION OF RESULTS
$CALCU %X=%NE(CCN(%O),0)
$EXIT  %X
$CALCU %Y=VOU(3) : %Z=VOU(4)
$USE   HEAD
$PRINT :
 'ABSOLUTE FREQUENCIES    F1 DUE TO CLASS MEANS X1  FROM  ' %3 :
 '---------------------   F2 DUE TO CLASS MEANS X2  FROM  ' %4 : :
 '            X1          F1          X2          F2'
$LOOK  KM1 KF1 KM2 KF2
$ENDMA
```

```
  H.BUSSE    11.1.82          BEISPIEL EINES GLIM-PROGRAMMS
$C
$C
$MACRO CLAF
$C      C L A F   CLASSIFICATION OF A VECTOR
$C       %1  INPUT VECTOR OF DIMENSION %6
$C       %2  %7 CLASS MEANS
$C       %3  %7 CLASS FREQUENCIES
$C       %4  %6 INPUT VALUES TRANSFORMED INTO CLASS MEANS
$C       %5  %6 INPUT VALUES TRANSFORMED INTO RANKS WITHIN CLASSES
$C       %6  DIMENSION OF INPUT VECTOR AND VECTORS %4 AND %5
$C       %7  NUMBER OF CLASSES REQUESTED
$C
$C   SORT INPUT %1 INTO TYM1 IN ASCENDENT ORDER
$DELET     TYM1 TYM2 TYM3
$CALCU %X=%7-1
$VARIA %6 TYM1
      : %7 TYM2
      : %X TYM3
$SORT      TYM1 %1
$CALCU     TYM3=%GL(%X,1)
$C
$C   COUNTING LIMITS %X AND %Y FOR EXTREME CLASSES 1 AND %7
$CALCU %Z=.5*%6/(%7+%SQRT(%7))
      : %Z=%Z*%LT(%Z,2)+2*%GT(%Z,2)
      : %Z=%TR(%Z+.9)
      : %X=TYM1(%Z+1)
      : %Y=TYM1(%6-%Z)
$C
$C   CLASS WIDTH %Z OF INTERIOR CLASSES ? UP TO %7-1
$CALCU %Z=(%Y-%X)/(%7-2)
$C
$C   TRANSFORMATION OF SORTED INPUT TYM1 INTO CLASS NUMBERS %4
$CALCU %4=(TYM1-%X)/%Z
      : %4=%TR(%IF(%GE(%4,0),%4,-1)+2)
      : %4=%LE(%4,1)+%GT(%4,1)*%4
      : %4=%IF(%GE(TYM1,%Y),%7,%4)
$C
$C   CLASS FREQUENCIES %3
$CALCU %3=0 : %3(%4)=%3(%4)+1
$C
$C   CLASS 'MEANS' %2
$CALCU %2     =%X+%Z*(%GL(%7,1)-1.5)
      : %2( 1)=%2( 1)-%Z
      : %2( 1)=%IF(%GT(TYM1( 1),%2( 1)),TYM1( 1),%2( 1))
      : %2(%7)=%2(%7)+%Z
      : %2(%7)=%IF(%LT(TYM1(%6),%2(%7)),TYM1(%6),%2(%7))
$C
$C   RANKING OF CLASSIFIED INPUT %4 WITHIN CLASSES IN ORDER
$C   TO GET ORDINATES %5 FOR THE HISTOGRAM
$CALCU TYM2(1      )=0
      : TYM2(1+TYM3)=%CU(%3(TYM3))
$CALCU %5=%GL(%6,1)-TYM2(%4)
$C
$C   TRANSFORMATION OF CLASSIFIED INPUT %4
$C   INTO CLASS MEANS %4
$CALCU TYM1=%4
      :    %4=%2(TYM1)
$C
$DELET TYM1 TYM2 TYM3
$ENDMA
```

```
  H.BUSSE    11.1.82          BEISPIEL FINES GLIM-PROGRAMMS
$C
$C
$MACRO MIMA
$C      M I M A    MINIMUM MEAN AND MAXIMUM OF A VECTOR
$C       %1  ARGUMENT
$C       %2  EXTREMA
$CALCU %2(1)=%1(1)
      : %2(1)=%IF(%LT(%1,%2(1)),%1,%2(1))
$CALCU %2(3)=%CU(%GE(%1,%2(1)))
      : %2(2)=%CU(%1)
      : %2(2)=%2(2)/%2(3)
$CALCU %2(3)=%1(1)
      : %2(3)=%IF(%GT(%1,%2(3)),%1,%2(3))
$ENDMA
$C
$C
$MACRO HEAD
$C      H E A D
$USE    SOCH
$PRINT /
  *'PROBIT ANALYSIS AND RELATED TECHNIQUES' *4 %Q
  *       BASED ON' *1 %G '  SIMULATED BINOMIAL RANDOM SAMPLES'
: *'-----------------------------------'
  *       USING GLIM-PROGRAM  DEBS/1' *1 CSR
$PRINT
: *   LINK FUNCTION Y(P) OF DATA GENERATION = ' %1
  *   Y = A+B*K = (X-M)/S+YO    A =' %A
  *     B =' %B
 :*   LINK FUNCTION Y(P) OF DATA ANALYSIS    = ' %2
  *                                 M =' %M
  *     S =' %S :
$DELET TYM1
$CALCU TYM1=%GL(%NU,1)
$PRINT 'SUBGROUP       I ' *4 TYM1
      : 'SIZE         N(I)' *4 NN
      : 'REGRESSOR    X(I)' *4 XX

      : 'REGRESSAND   Y(I)' *4 YY
      : 'PROBABILITY  P(I)' *4 PP
$DELET TYM1
$ENDMA
$C
$C
$MACRO SOCH
$C      S O C H    SET OUTPUT CHANNEL WITH FORMAT
$C      %Y  WISHED COLUMN NUMBER WIDTH  FOR GRAPHICAL OUTPUT
$C      %Z  WISHED    ROW NUMBER HEIGHT FOR GRAPHICAL OUTPUT
$CALCU %X=VOU(3)-15
      : %Y=%Y*%LE(%Y,%X)+%X*%GT(%Y,%X)
      : %Y=%Y*%GE(%Y,112)+112*%LT(%Y,112)
$CALCU %X=VOU(4)
      : %Z=%Z*%LE(%Z,%X)+%X*%GT(%Z,%X)
      : %Z=%Z*%GE(%Z,10)+10*%LT(%Z,10)
$CALCU %X=VOU(2)
$OUTPU %X %Y %Z
$ENDMA
$C
$C
$MACRO BLAP
$C      B L A P    PRINT %X BLANK LINES. AT LEAST ONE
$PRINT ' '
$CALCU %X=(%X-1)*%GT(%X,1)
$ENDMA
$C
```

```
$C
$C
$MACRO GRAP
$C      G R A P   TWO-DIMENSIONAL PRESENTATION
$C       %1  VECTOR OF %5 ORDINATES
$C       %2  VECTOR OF %G ABSCISSA REFERRED TO %1
$C       %3  TEXT FOR ORDINATE
$C       %4  TEXT FOR ABSCISSA
$C       %5  MACRO WITH LETTER OF ORDINATE VECTOR
$C           ACTUALLY TO BE PLOTTED
$C
$CALCU %Y=VOU(3) : %Z=VOU(4)
$USE   HEAD
$PRINT ' '
$DELET ?%5
$CALCU ?%5=%1
$PLOT  ?%5 %2
$DELET ?%5
$PRINT : 'ORDINATE : ' %3
       : '                                                    '
         'ABSCISSA : ' %4 :
$ENDMA
$C
$C ..........................................................
$C
$MACRO HIS1
$C      H I S 1   HISTOGRAM WITH ONE DISTRIBUTION ONLY
$C       %1  ORDINATE VECTOR CONTAINING %G RANKS WITHIN CLASSES
$C       %2  ABSCISSA VECTOR CONTAINING %G CLASS MEANS
$C       %3  TEXT
$C       %4  MINIMUM MEAN AND MAXIMUM OF ORIGINARY ABSCISSA-VARIABLE
$C       %5  MACRO WITH LETTER OF ORDINATE VECTOR ACTUALLY TO BE PLOTTED
$C
$CALCU %Y=VOU(3)
     : %Z=0 : %Z=%IF(%GT(%1,%Z),%1,%Z) : %Z=%Z+4
     : VOU(5)=VOU(4)-%Z
$USE   HEAD
$CALCU %X=VOU(5)*%GT(VOU(5),0)
$WHILE %X BLAP
$C
$C  PLOT
$DELET ?%5
$CALCU ?%5=%1
$PLOT  ?%5 %2
$DELET ?%5
$C
$C  LEGENDA
$CALCU %X=%4(1) : %Y=%4(2) : %Z=%4(3)
$PRINT : *1 %H '  ABSOLUTE FREQUENCIES DUE TO ' %3
                '   MINIMUM =' *4 %X :
   '                                        ' BLAT
                '      MEAN =' %Y :
   '                                        ' BLAT
                '   MAXIMUM =' %Z
$ENDMA
```

```
  H.BUSSE    11.1.82          BEISPIEL EINES GLIM-PROGRAMMS
$C
$C
$MACRO HIS2
$C      H I S 2    HISTOGRAM CONTAINING TWO DISTRIBUTIONS
$C       %1  FIRST  VECTOR WHICH FREQUENCIES ARE TO BE SHOWN
$C       %2  SECOND VECTOR WHICH FREQUENCIES ARE TO BE SHOWN
$C       %3  TEXT DUE TO %1
$C       %4  TEXT DUE TO %2
$C       %5  MINIMA MEANS AND MAXIMA
$C            1 MIN(%1)
$C            2 MEA(%1)
$C            3 MAX(%1)
$C            4 MIN(%2)
$C            5 MEA(%2)
$C            6 MAX(%2)
$C       %6  MACRO WITH LETTER OF ORDINATE VECTOR ACTUALLY TO BE PLOTTED
$C
$C  DEFINE VECTORS OF DOUBLE SIMULATION LENGTH %K
$C              AND OF DOUBLE CLASS NUMBER      %N
$DELET    TYM1 CZA1 CZA2 CZA3 SOW1 SOW2
$CALCU %K=2*%G
$C  CONDITIONAL NUMBER %H OF FREQUENCY CLASSES
$CALCU %H=2*CCF(%O)
     : %H=%IF(%GT(%H,6),%H,6)
     : %X=VOU(3)-25
     : %H=%IF(%LT(%H,%X),%H,%X)
$VARIA %G TYM1
     : %K CZA1 CZA2 CZA3
     : %H SOW1 SOW2
$C
$C  COMBINE INPUT VECTOR FOR CLASSIFICATION
$CALCU TYM1=%GL(%G,1)
     : CZA1(TYM1   )=%1
     : CZA1(TYM1+%G)=%2
$C  CLASSIFY COMBINED INPUT VECTOR
$ARGUM CLAF CZA1 SOW1 SOW2 CZA2 CZA3
            %K   %H
$USE   CLAF
$C
$C  HEADING AND DOUBLE HISTOGRAM
$CALCU %Y=VOU(3)
     : %Z=0 : %Z=%IF(%GT(CZA3,%Z),CZA3,%Z) : %Z=%Z+4
     : VOU(5)=VOU(4)-%Z
$USE   HEAD
$CALCU %X=VOU(5)*%GT(VOU(5),0)
$WHILE %X BLAH
$C
$C  PLOT
$DELET ?%6
$CALCU ?%6=CZA3
$PLOT  ?%6 CZA2
$DELET ?%6
$C
$C  LEGENDA
$PRINT : *1 %H ' FREQUENCIES OF TWICE' %G ' RESULTS'
       : '      ' BLAT '      MIN     MEAN      MAX'
$CALCU %X=%5(1) : %Y=%5(2) : %Z=%5(3)
$PRINT 'FROM ' %3 *4 %X %Y %Z
$CALCU %X=%5(4) : %Y=%5(5) : %Z=%5(6)
$PRINT ' AND ' %4 *4 %X %Y %Z
$C
$DELET TYM1 CZA1 CZA2 CZA3 SOW1 SOW2
$ENDMA
```

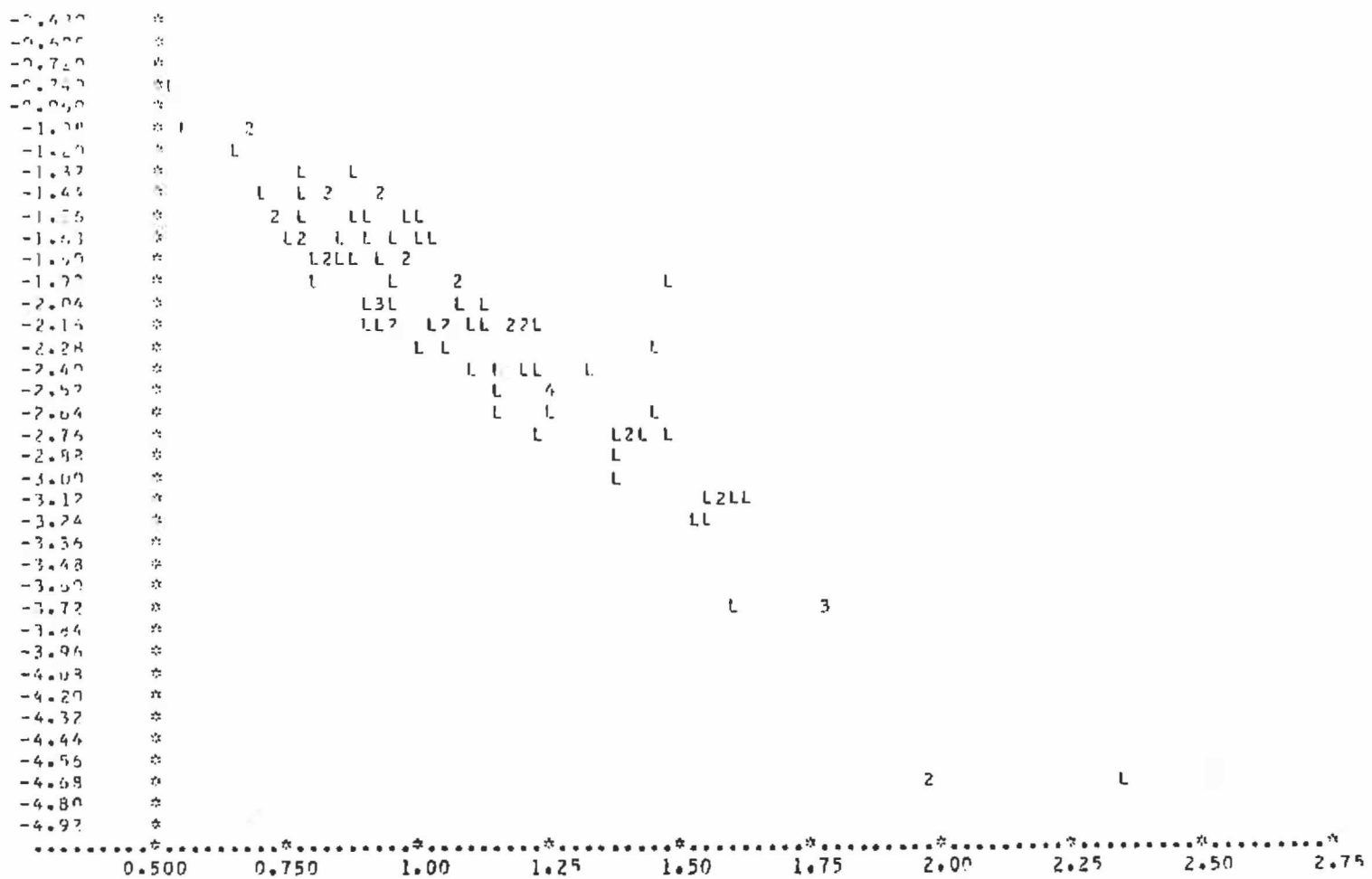
```
  H.BUSSE    11.1.82          BEISPIEL EINES GLIM-PROGRAMMS
$C
$C
$C   1. EXAMPLE: LIKE PREVIOUSLY DEFINED
$C
$SUBFILE EXA1
$UNITS 5
$CALCU %G=100
      : %Q=1
$DELET NN PP
$CALCU NN=PP=0
$EDIT  NN   10  10  10  10  10
     :  PP 0.15 .35 .55 .75 .95
$CALCU %I=1 : %J=1
      : %M=2 : %S=1
$EDIT  CSR 8 12 31
     :  CCB 1  1  1
     :  CCF 0  0 15
     :  CCH 0  3  4
     :  CCN 1  1  0
     :  VOU 20 6 112 41 0
$RETURN
$C
$C ------------------------------------------------
$C
$C    2. EXAMPLE:
$C
$SUBFILE EXA2
$UNITS 6
$CALCU %G=100
      : %Q=2
$DELET NN PP
$CALCU NN=PP=0
$EDIT  NN    7   8   9  11   8   7
     :  PP 0.10 .20 .30 .50 .75 .95
$CALCU %I=2 : %J=1
      : %M=2 : %S=1
$EDIT  CSR 11 10 42
     :  CCB 1  1  1
     :  CCF 0  0 17
     :  CCH 0  3  4
     :  CCN 1  1  0
     :  VOU 20 6 112 41 0
$RETURN
$C
$FINISH
```

PROBIT ANALYSIS AND RELATED TECHNIQUES 1.000

BASED ON 100. SIMULATED BINOMIAL RANDOM SAMPLES
USING GLIM-PROGRAM (V3.77) 8. 17. 31.

LINK FUNCTION TYPE OF DATA GENERATION = PROBIT Y = A+BX = (X-1)/3+Y0 A = -2.000 B = 1.000
LINK FUNCTION TYPE OF DATA ANALYSIS = PROBIT A = 2.000 S = 1.000

DIMENSION		I	1.000	2.000	3.000	4.000	5.000
SIZE		N(I)	10.00	10.00	10.00	10.00	10.00
DOSAGE		X(I)	0.6065	1.366	2.125	2.885	3.645
REGRESSAND		Y(I)	-1.394	-0.6339	0.1257	0.8853	1.645
PROBABILITY		P(I)	0.0817	0.2631	0.5500	0.8120	0.9500



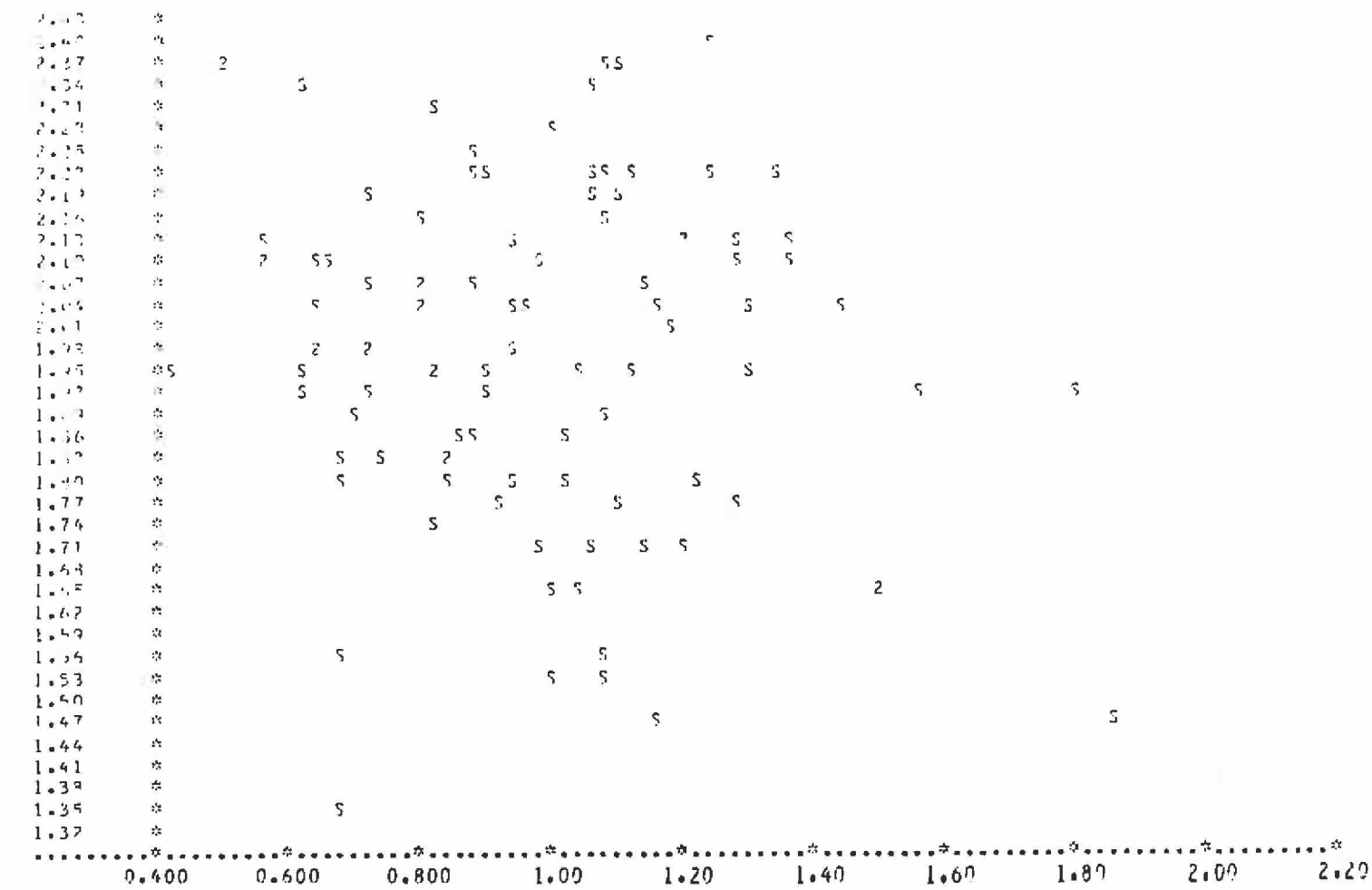
ORDINATE : A-COEFFICIENT OF Y=A+B*X

ABSCISSA : B-COEFFICIENT OF Y=A+B*X

PROBIT ANALYSIS AND RELATED TECHNIQUES 1.000 BASED ON 100. SIMULATED PSEUDO-REAL RANDOM SAMPLES
USING GLIM-PROGRAM (1982) 3. 12. 81.

LINK FUNCTION Y(P) OF DATA GENERATION = PROBIT Y = A+B*X = (X-M)/S+YC A = -2.000 B = 1.000
LINK FUNCTION Y(P) OF DATA ANALYSIS = PROBIT A = -2.000 B = 1.000

[illegible]	I	1.000	2.000	3.000	4.000	5.000
SIZE	N(I)	10.00	10.00	10.00	10.00	10.00
[illegible]	X(I)	0.6066	1.386	1.820	2.386	3.465
[illegible]	Y(I)	-1.394	-0.6339	0.1257	0.8953	1.645
PROBABILITY	P(I)	0.0317	0.2631	0.5500	0.8120	0.9500



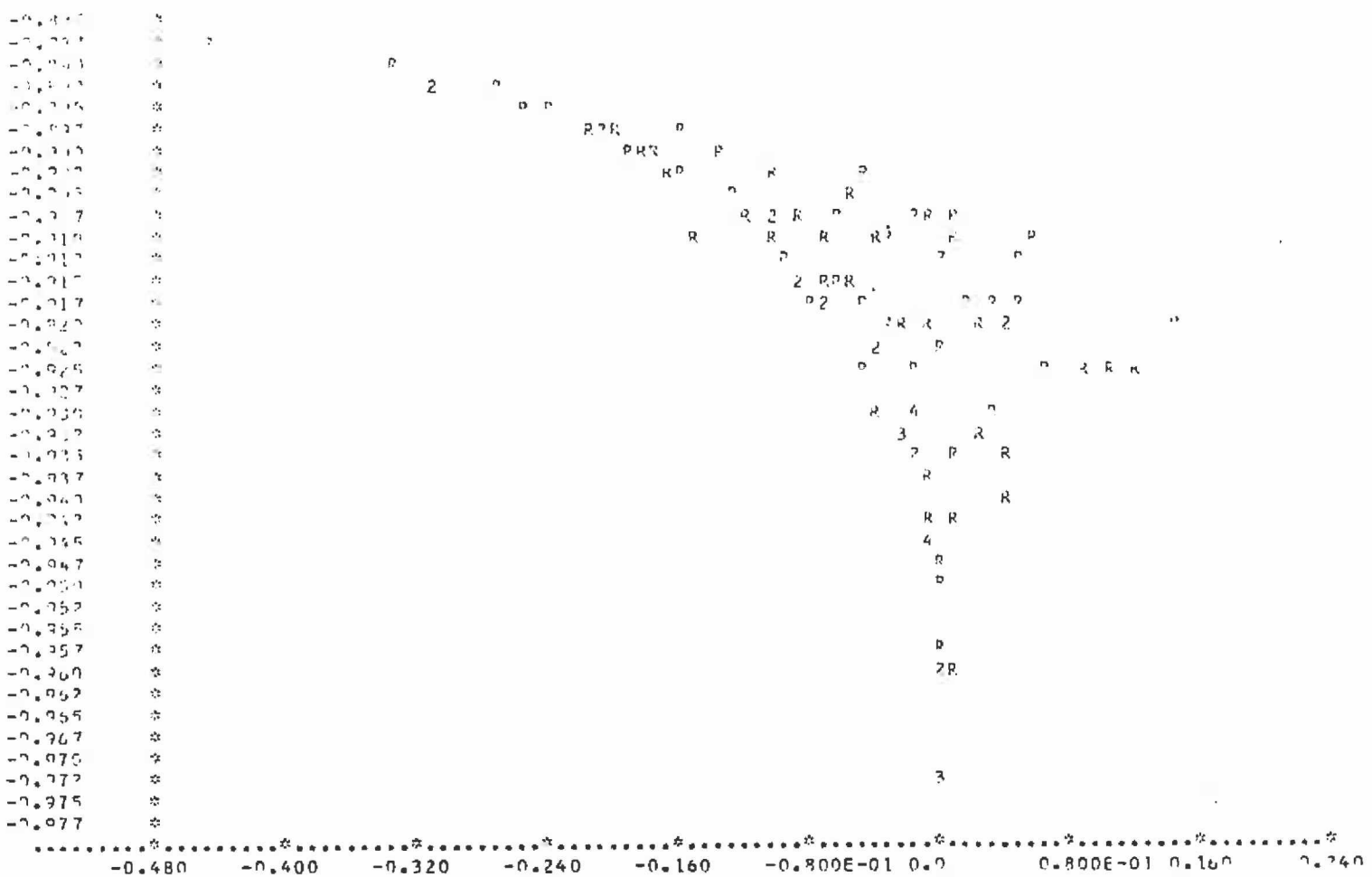
ORDINATE : M-COEFFICIENT OF Y=(X-M)/S

ABSCISSA : S-COEFFICIENT OF Y=(X-M)/S

PROBIT ANALYSIS AND RELATED TECHNIQUES 1.000 BASED ON 100 SIMULATED BINOMIAL RANDOM SAMPLES
USING GLIM-[illegible] [illegible] 4. 14. 31.

LINK FUNCTION Y(P) OF DATA GENERATION = PROBIT Y = A+B*X = (X-M)/S+Y0 A = -2.000 S = 1.000
LINK FUNCTION Y(P) OF DATA ANALYSIS = PROBIT M = 2.000 S = 1.000

[illegible]	I	1.000	2.000	3.000	4.000	5.000
SIZE	N(I)	10.00	10.00	10.00	10.00	10.00
[illegible]	Y(I)	0.6795	1.366	2.134	2.795	3.665
[illegible]	Y(I)	-1.396	-0.6339	0.1257	0.8053	1.643
PROBABILITY	P(I)	0.0813	0.2631	0.5500	0.6120	0.9500



ORDINATE : CORRELATION(A,B) IN Y=A+B*X

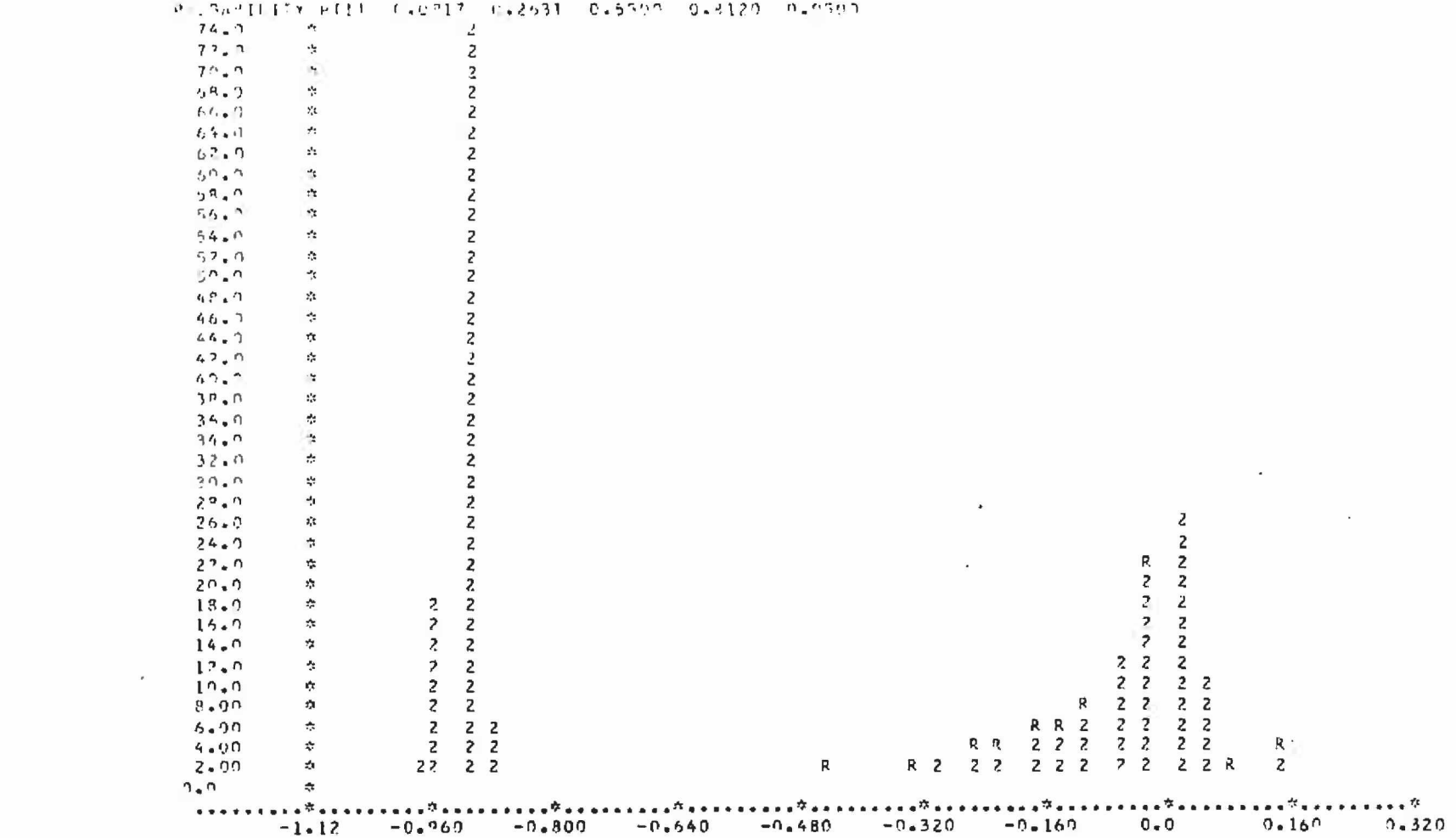
ABSCISSA : CORRELATION(M,S) IN Y=(X-M)/S

PROBIT ANALYSIS AND RELATED TECHNIQUES [illegible]

BASED ON 100. SIMULATED [illegible]
USING CLIP-PROGRAM [illegible]

LIN. FUNCTION Y(X) OF DATA GENERATION = PROBIT Y = A+B*X = (X-M)/S+5 A = -3.000 B = 1.000
LIN. FUNCTION Y(P) OF DATA ANALYSIS = PROBIT A = 3.000 B = 1.000

	I	1.000	2.000	3.000	4.000	5.000
[illegible]	N(I)	10.00	10.00	10.00	10.00	10.00
[illegible]	X(I)	0.8965	1.266	2.124	2.785	3.643
[illegible]	Y(I)	-1.394	-0.6339	0.1257	0.8052	1.645
[illegible] PROBABILITY	P(I)	0.0917	0.2631	0.5500	0.8120	0.9503



30. FREQUENCIES OF TWICE 100. RESULTS

	MIN	MEAN	MAX
FROM CORRELATION(A,B) IN Y=A+B*X	-0.9735	-0.9203	-0.8865
AND CORRELATION(M,S) IN Y=(X-M)/S	-0.4472	-0.0561	0.1466

PROBIT ANALYSIS AND RELATED TECHNIQUES 1.400 BASED ON 100. SIMULATED BINOMIAL RANDOM SAMPLES
USING GLIM-PROGRAM DOSIS 9. 12. 31.

LINK FUNCTION Y(P) OF DATA GENERATION = PROBIT Y = A+B*X = (X-1)/3+X0 A = -2.000 B = 1.000
LINK FUNCTION Y(P) OF DATA ANALYSIS = PROBIT A = 2.000 B = 1.000

[illegible]	I	1.000	2.000	3.000	4.000	5.000
SIZE	N(I)	10.00	10.00	10.00	10.00	10.00
MEAN DOSE	X(I)	0.6065	1.366	2.126	2.885	3.645
CORRESPOND.	Y(I)	-1.394	-0.6339	0.1257	0.8853	1.645
PROBABILITY	P(I)	0.0817	0.2631	0.5398	0.8120	0.9500

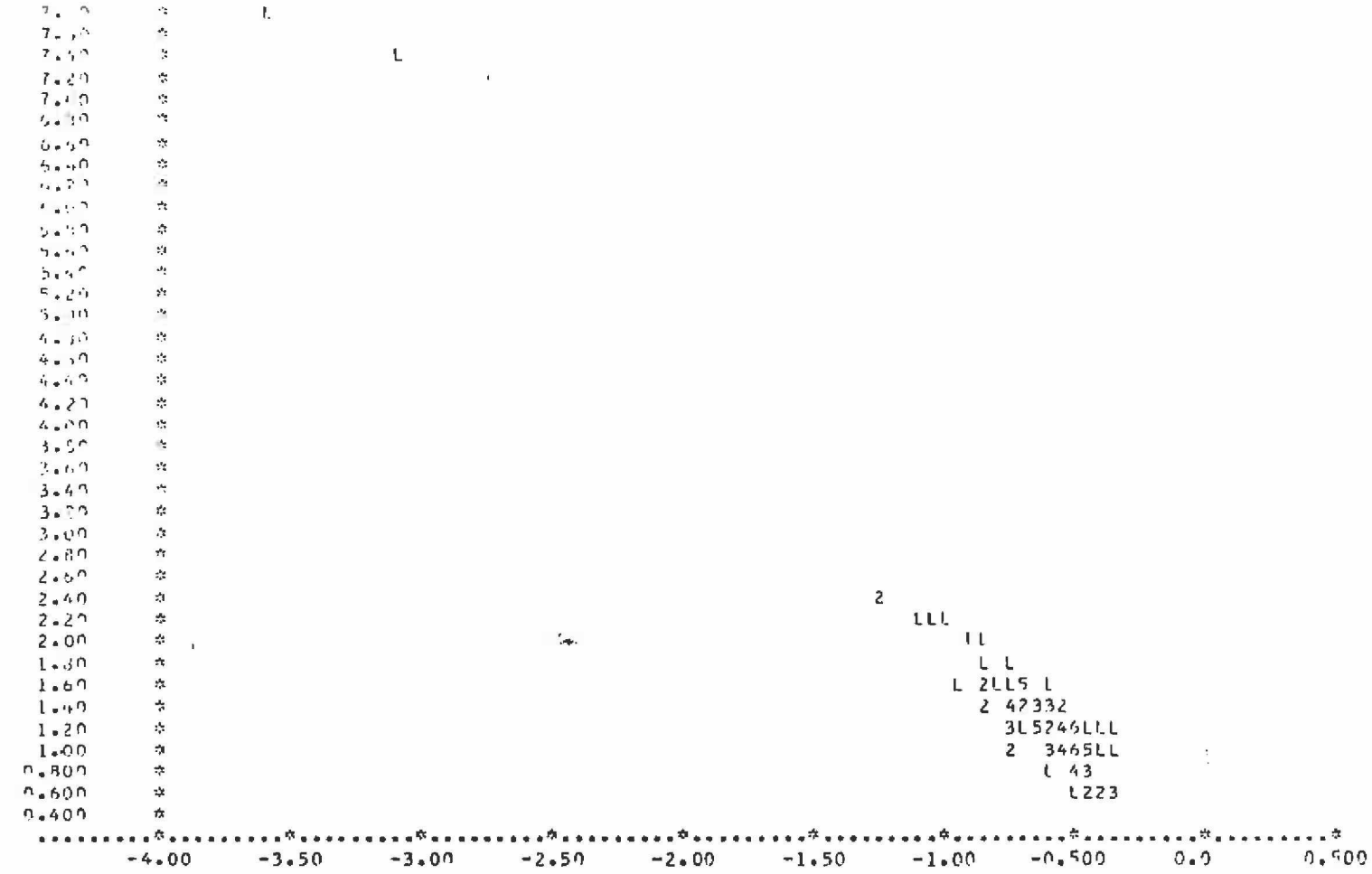
ABSOLUTE FREQUENCIES F1 DUE TO CLASS MEANS X1 FROM CORRELATION(1,2) IN Y=A+B*X
F2 DUE TO CLASS MEANS X2 FROM CORRELATION(1,2) IN Y=(X-B)/A

	X1	F1	X2	F2
1	-0.2735	2.000	-0.3546	2.000
2	-0.2674	1.000	-0.2938	2.000
3	-0.1633	1.000	-0.2611	2.000
4	-0.2571	3.000	-0.2273	2.000
5	-0.1898	2.000	-0.1975	4.000
6	-0.1447	6.000	-0.1656	6.000
7	-0.0976	2.000	-0.1338	3.000
8	-0.7324	12.00	-0.1070	6.000
9	-0.1242	9.000	-0.7012E-01	9.000
10	-0.1200	13.00	-0.3824E-01	15.00
11	-0.1138	12.00	-0.6444E-02	31.00
12	-0.0977	16.00	0.2533E-01	9.000
13	-0.0815	10.00	0.5723E-01	6.000
14	-0.0753	6.000	0.9936E-01	1.000
15	-0.0745	4.000	0.1466	3.000

PROBIT ANALYSIS AND RELATED TECHNIQUES 3.000 BASED ON 100. SIMULATED BINOMIAL RANDOM SAMPLES
USING GLIM-PROGRAM VERSION 11. 10. 42.

LINK FUNCTION Y(P) OF DATA GENERATION = LOGIT Y = A+B*X = (X-M)/S+Y0 A = -2.000 L = 1.000
LINK FUNCTION Y(P) OF DATA ANALYSIS = PROBIT M = 3.000 S = 1.000

SUBGROUP	I	1.000	2.000	3.000	4.000	5.000	6.000
SIZE	N(I)	7.000	9.000	7.000	11.00	9.000	7.000
OF X-VALUE	X(I)	-1.375	-0.1112	1.153	2.417	3.681	4.944
LOGISTIC	Y(I)	-3.375	-2.111	-0.8473	0.4166	1.681	2.944
PROBABILITY	P(I)	0.9668	0.8920	0.7000	0.3972	0.1570	0.0501



ORDINATE : A-COEFFICIENT OF Y=A+B*X

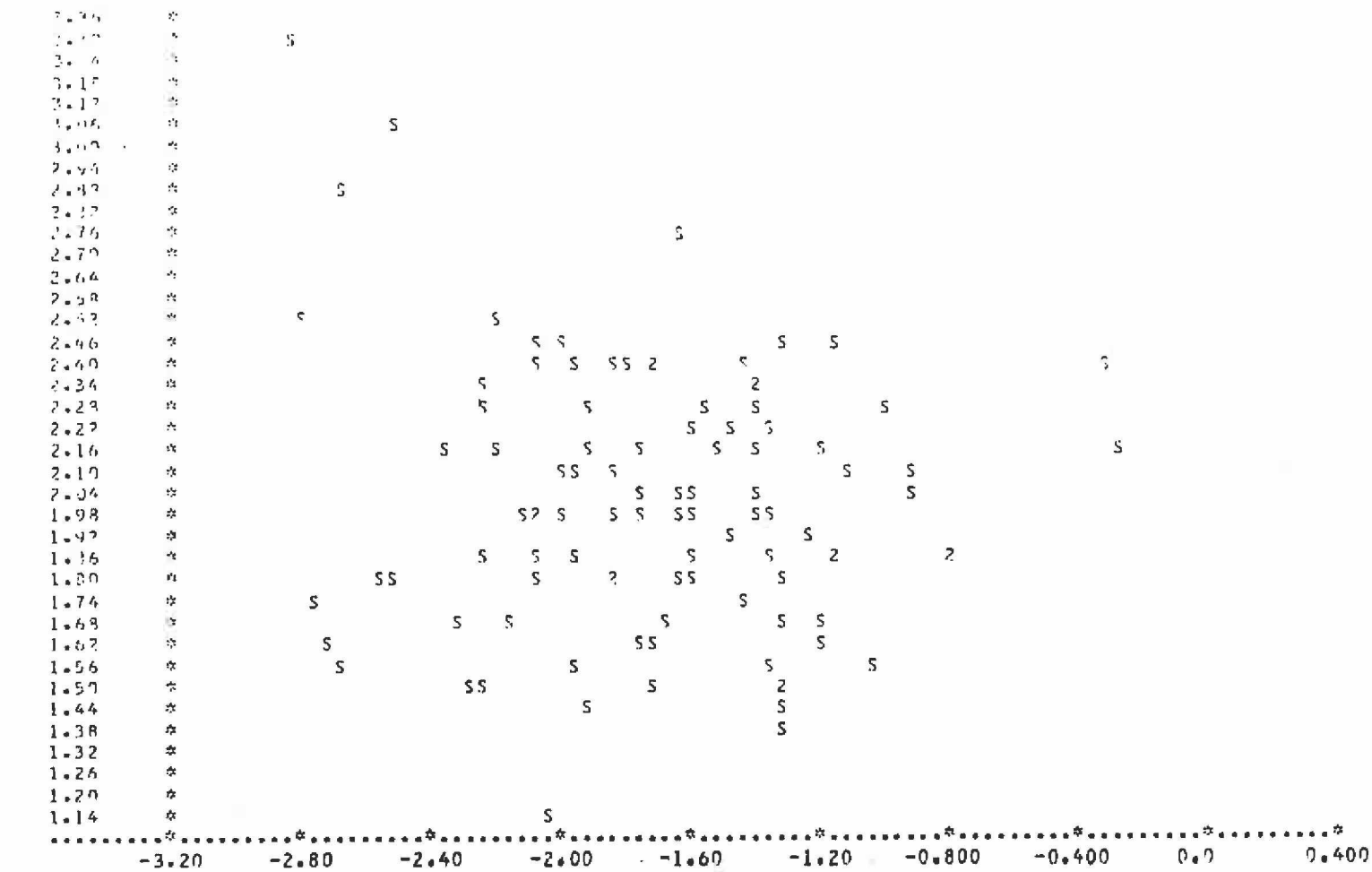
ABSCISSA : B-COEFFICIENT OF Y=A+B*X

PROBIT ANALYSIS AND RELATED TECHNIQUES 2.010

ASSAY OF 100. SIMULATED BINOMIAL RANDOM SAMPLES
UCLA BIO-PROGRAM 9 MAY 71 11. 10. 42.

LOW FUNCTION Y(P) OF DOSE PENETRATION = LOGIT Y = A+BX = (X-M)/S+XO A = -2.305 r = 1.000
LOW FUNCTION Y(P) OF DOSE ANALYSIS = PROBIT A = -2.021 S = 1.000

SUBGROUP	I	1.000	2.000	3.000	4.000	5.000	6.000
SIZE	N(I)	7.000	1.000	9.000	11.00	9.000	7.000
RESPONSE	X(I)	-1.373	-1.111	-.1143	.2417	.6481	1.264
TRANSFORM	Y(I)	-2.375	-2.111	-1.843	-1.606	-1.481	-1.264
PROBABILITY	P(I)	.3568	.1020	.3778	.5173	.6778	.8980



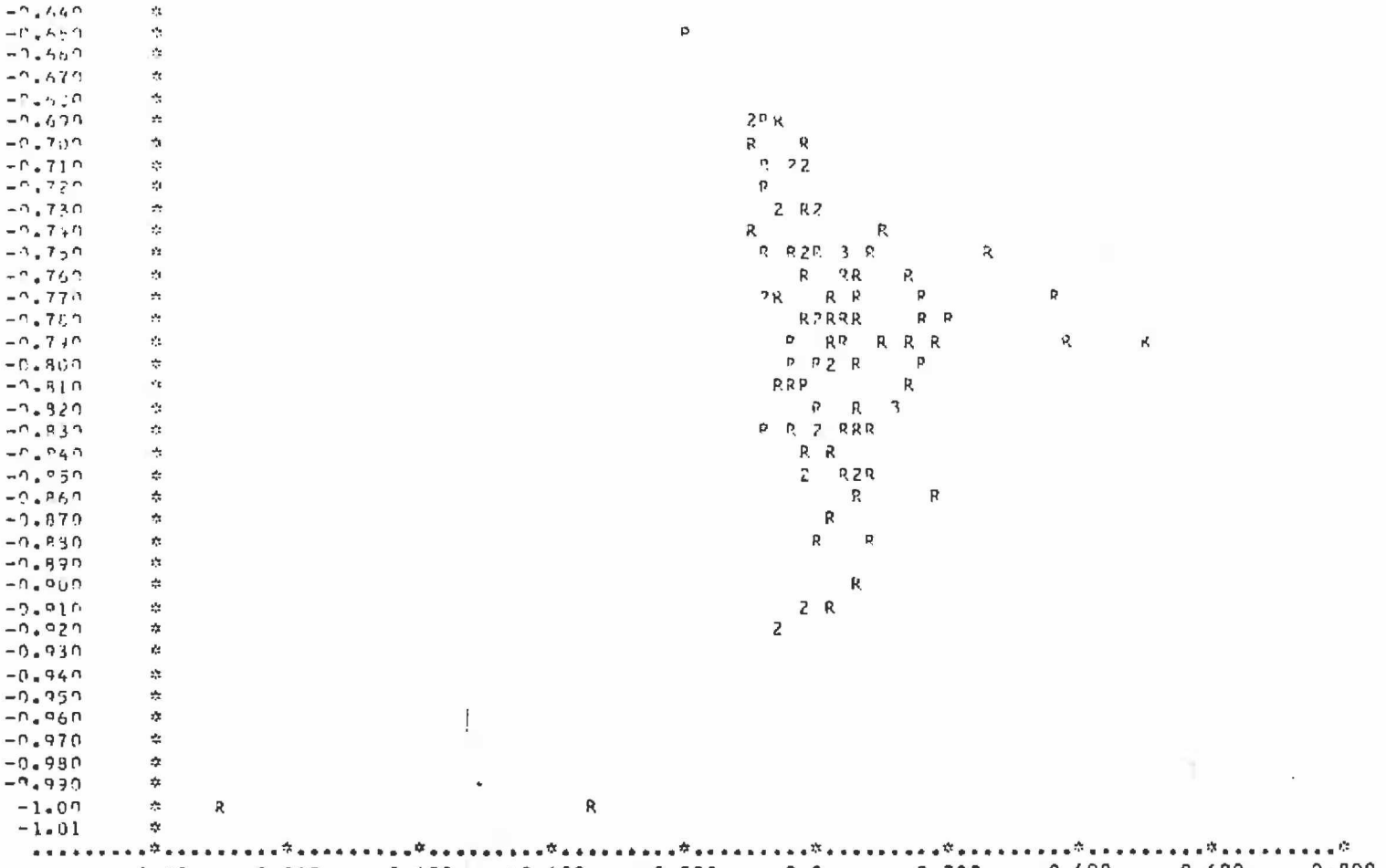
ORDINATE : M-COEFFICIENT OF Y=(X-M)/S

ABSCISSA : S-COEFFICIENT OF Y=(X-M)/S

PROBIT ANALYSIS AND RELATED TECHNIQUES 2.000 BASED ON 100. SIMULATED BINOMIAL RANDOM SAMPLES
USING GLIM-PROGRAM PROB/1 11. 10. 42.

LINK FUNCTION Y(P) OF DATA GENERATION = LOGIT Y = A+BX = (X-M)/S+XI M = -2.000 S = 1.000
LINK FUNCTION Y(P) OF DATA ANALYSIS = PROBIT M = 2.000 S = 1.000

SUBGROUP	I	1.000	2.000	3.000	4.000	5.000	6.000
SIZE	N(I)	7.000	8.000	9.000	11.00	8.000	7.000
PROGRESSOR	X(I)	-1.375	-0.1112	1.153	2.417	3.681	4.944
PROGRESSION	Y(I)	-3.375	-2.111	-0.8473	0.4166.	1.681	2.944
PROBABILITY	P(I)	0.9669	0.8920	0.7900	0.3773	0.1570	0.0500



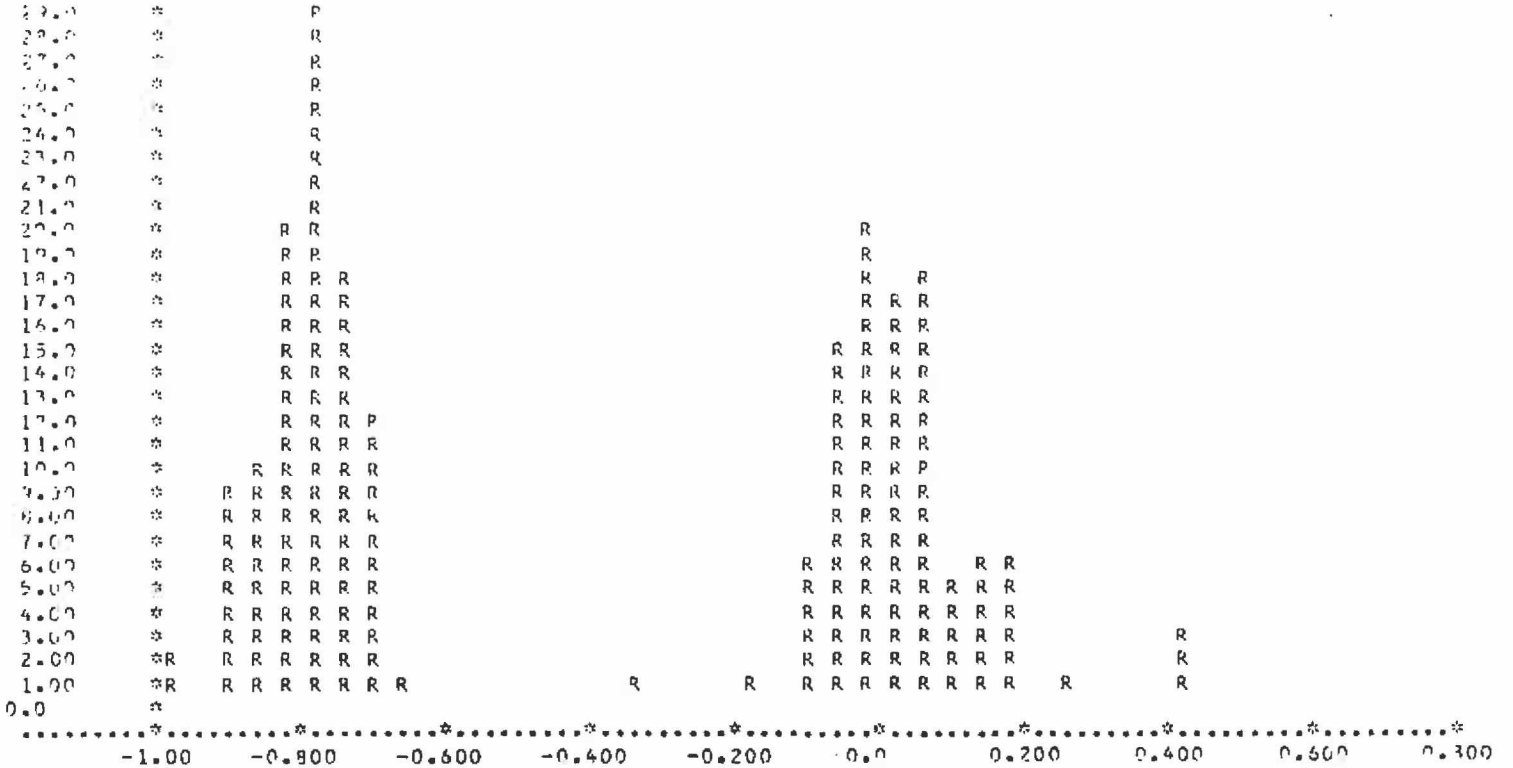
ORDINATE : CORRELATION(A,B) IN Y=A+B*X

ABSCISSA : CORRELATION(M,S) IN Y=(X-M)/S

PROBIT ANALYSIS AND RELATED TECHNIQUES [illegible] BASED ON 100. SIMULATED BINOMIAL RANDOM SAMPLES
USING GLIM-PROGRAM DEZEMBER 11. 10. 42.

LINK FUNCTION Y(P) OF DATA GENERATION = LOGIT Y = A+B*X = (X-M)/S A = -3.000 B = 1.000
LINK FUNCTION Y(P) OF DATA ANALYSIS = PROBIT M = 3.000 S = 1.000

[illegible]	I	1.000	2.000	3.000	4.000	5.000	6.000
SIZE	N(I)	7.000	9.000	11.00	11.00	9.000	7.000
[illegible]	X(I)	-1.375	-0.1112	1.157	2.417	3.681	4.944
[illegible]	Y(I)	-3.375	-2.111	-0.8473	0.4156	1.681	2.944
PROPORTION	P(I)	0.9667	0.8920	0.7909	0.2973	0.1970	0.0500



34. FREQUENCIES OF TWICE 100. RESULTS

	MIN	MEAN	MAX
FROM CORRELATION(A,B) IN Y=A+B*X	-0.9992	-0.7922	-0.6548
AND CORRELATION(M,S) IN Y=(X-M)/S	-0.9004	0.0152	0.4919

PROBIT ANALYSIS AND RELATED TECHNIQUES 2.000

BASED ON 100. SIMULATED BINOMIAL RANDOM SAMPLES
USING GLIM-PROGRAM DEBS/1 11. 10. 42.

LINK FUNCTION Y(P) OF DATA GENERATION = LOGIT Y = A+B*X = (X-M)/S+Y0 A = -2.000 B = 1.000
LINK FUNCTION Y(P) OF DATA ANALYSIS = PROBIT M = 2.000 S = 1.000

SUBGROUP	I	1.000	2.000	3.000	4.000	5.000	6.000
SIZE	N(I)	7.000	8.000	9.000	11.00	8.000	7.000
REGRESSOR	X(I)	-1.375	-0.1112	1.153	2.417	3.681	4.944
REGRESSAND	Y(I)	-3.375	-2.111	-0.8473	0.4166	1.681	2.944
PROBABILITY	P(I)	0.9669	0.8920	0.7000	0.3973	0.1570	0.0500

ABSOLUTE FREQUENCIES F1 DUE TO CLASS MEANS X1 FROM CORRELATION(A,B) IN Y=A+B*X
F2 DUE TO CLASS MEANS X2 FROM CORRELATION(M,S) IN Y=(X-M)/S

	X1	F1	X2	F2
1	-0.9387	2.000	-0.2473	2.000
2	-0.9087	5.000	-0.1745	1.000
3	-0.8937	1.000	-0.1381	0.0
4	-0.8787	2.000	-0.1017	6.000
5	-0.8636	3.000	-0.6534E-01	13.00
6	-0.8486	6.000	-0.2895E-01	20.00
7	-0.8336	8.000	0.7440E-02	16.00
8	-0.8186	7.000	0.4383E-01	15.00
9	-0.8036	9.000	0.8022E-01	9.000
10	-0.7886	11.00	0.1166	7.000
11	-0.7736	12.00	0.1530	4.000
12	-0.7585	6.000	0.1894	3.000
13	-0.7435	10.00	0.2258	0.0
14	-0.7285	5.000	0.2622	1.000
15	-0.7135	6.000	0.2986	0.0
16	-0.6985	4.000	0.3350	0.0
17	-0.6685	3.000	0.4077	3.000

```
$C
$C
$C    ***************************
$C    *                         *     USER-DEFINED FUNCTIONS
$C    *  G L I M  -  O W L I    *
$C    *                         *     FOR LINK AND ERROR
$C    ***************************
$C
$C    15.3.1982  H.BUSSE
$C
$C    DEFINITIONS FOR LINK FUNCTIONS
$C    ------------------------------
$C

$C     O W N I    LINEAR PREDICTORS %LP AS LINK FUNCTIONS
$C                OF FITTED VALUES %FV
$C
$C     O W N 1    FITTED VALUES %FV AS INVERSE LINK FUNCTIONS
$C                OF LINEAR PREDICTORS %LP
$C
$C     O W N 2    DERIVATIVES %DR OF LINEAR PREDICTORS %LP BY FITTED VALUES %FV
$C

$C
$SUBFI LIDE                          !   IDENTITY
$MACRO OWNI $CALCU %LP=%FV           $ENDMA
$MACRO OWN1 $CALCU %FV=%LP           $ENDMA
$MACRO OWN2 $CALCU %DR= 1            $ENDMA
$RETURN
$C
$C
$SUBFI LREC                          !   RECIPROCAL
$MACRO OWNI $CALCU %LP=1/%FV         $ENDMA
$MACRO OWN1 $CALCU %FV=1/%LP         $ENDMA
$MACRO OWN2 $CALCU %DR=-%LP*LP       $ENDMA
$RETURN
$C
$C
$SUBFI LREN                          !   NEGATIVE RECIPROCAL
$MACRO OWNI $CALCU %LP=-1/%FV        $ENDMA
$MACRO OWN1 $CALCU %FV=-1/%LP        $ENDMA
$MACRO OWN2 $CALCU %DR=%LP*LP        $ENDMA
$RETURN
$C
$C
$SUBFI LOGA                          !   LOGARITHM
$MACRO OWNI $CALCU %LP=%LOG(%FV)     $ENDMA
$MACRO OWN1 $CALCU %FV=%EXP(%LP)     $ENDMA
$MACRO OWN2 $CALCU %DR=1/%FV         $ENDMA
$RETURN
$C
```

```
$C
$C    LINK FUNCTIONS FOR BINOMIAL DISTRIBUTION
$C    ----------------------------------------
$C

$SUBFI LIDB                                        !  IDENTITY
$MACRO OWN1 $CALCU %LP=%FV/BD                      $ENDMA
$MACRO OWN1 $CALCU %FV=%LP*BD                      $ENDMA
$MACRO OWN2 $CALCU %DR=1/BD                        $ENDMA
$RETURN
$C
$C
$SUBFI LOGB                                        !  LOGARITHM
$MACRO OWN1 $CALCU %LP=%LOG(%FV/BD)                $ENDMA
$MACRO OWN1 $CALCU %FV=%EXP(%LOG(BD)+%LP)          $ENDMA
$MACRO OWN2 $CALCU %DR=1/%FV                       $ENDMA
$RETURN
$C
$C
$SUBFI LOGI                                              !   LOGIT
$MACRO OWN1 $CALCU %LP=%FV/BD
                 : %LP=%LOG(%LP/(1-%LP))                 $ENDMA
$MACRO OWN1 $CALCU %FV=BD/(%EXP(-%LP)+1)                 $ENDMA
$MACRO OWN2 $CALCU %DR=BD/((BD-%FV)*%FV)                 $ENDMA
$RETURN
$C
$C
$SUBFI LPRO                                              !   PROBIT
$MACRO OWN1 $CALCU %LP=%ND(%FV/BD)                       $ENDMA
$MACRO OWN1 $CALCU %FV=BD*%NP(%LP)                       $ENDMA
$MACRO OWN2 $CALCU %DR=%EXP(%LP*%LP/2)
                      *%SQRT(2*%PI)/BD                   $ENDMA
$RETURN
$C
$C
$SUBFI LLOG                                              !   LOGLOG
$MACRO OWN1 $CALCU %LP=-%LOG(-%LOG( %FV/BD))             $ENDMA
$MACRO OWN1 $CALCU %FV=BD*%EXP(-%EXP(-%LP))              $ENDMA
$MACRO OWN2 $CALCU %DR=-1/(%FV*%LOG(%FV/BD))             $ENDMA
$RETURN
$C
$C
$SUBFI LCLO                                              !   CLOGLOG
$MACRO OWN1 $CALCU %LP=%LOG(-%LOG(1-%FV/BD)              $ENDMA
$MACRO OWN1 $CALCU %FV=BD*(1-%EXP(-%EXP(%LP)))           $ENDMA
$MACRO OWN2 $CALCU %DR=-1/((BD-%FV)*%LOG(1-%FV/BD))      $ENDMA
$RETURN
$C
```

```
$C
$C    DEFINITIONS FOR DENSITY FUNCTION
$C    --------------------------------
$C
$C     O W N O     INITIAL ESTIMATES FOR LINEAR PREDICTORS %LP
$C
$C     O W N 3     VARIANCE FUNCTIONS %VA AS SECOND DERIVATIVES
$C                 OF B-FUNCTION BY ITS NATURAL PARAMETER T
$C
$C     O W N 4     INCREMENTS %DI OF DEVIANCE %DV
$C

$C
$SUBFI ANOR  !  NORMAL DISTRIBUTION
$MACRO OWN0 $CALCU %FV=YV             $ENDMA
$MACRO OWN3 $CALCU %VA=1              $ENDMA
$MACRO OWN4 $CALCU %DI=(%YV-%FV)**2 $ENDMA
$RETURN
$C

$SUBFI AGAM  !  GAMMA DISTRIBUTION
$MACRO OWN0 $CALCU %FV=YV+.5                    $ENDMA
$MACRO OWN3 $CALCU %VA=%FV*%FV                  $ENDMA
$MACRO OWN4 $CALCU %DI=%IF(%EQ(%YV,0),1,%YV/%FV)
             : %DI=2*(%YV/%FV-%LOG(%DI)-1) $ENDMA
$RETURN
$C

$C
$SUBFI ABIN  !  BINOMIAL DISTRIBUTION
$MACRO OWN0 $CALCU %FV=%IF(%EQ(YV, 0),   .5, YV)
                  : %FV=%IF(%EQ(YV,BD),BD-.5,%FV)          $ENDMA
$MACRO OWN3 $CALCU %VA=%FV*(1-%FV/BD)                      $ENDMA
$MACRO OWN4 $CALCU %DI=2*(%YV*%LOG(%IF(%EQ(%YV,0),1,%YV/%FV))
                       +(BD-%YV)*%LOG(%IF(
                      %EQ(BD,%YV),1,(BD-%YV)/(BD-%FV))))$ENDMA
$RETURN
$C

$C
$SUBFI APOI  !  POISSON DISTRIBUTION
$MACRO OWN0 $CALCU %FV=YV+.5                                $ENDMA
$MACRO OWN3 $CALCU %VA=%FV                                  $ENDMA
$MACRO OWN4 $CALCU %DI=2*(%YV*%LOG(%IF(%EQ(%YV,%FV)
                                1,%YV/%FV))-%YV+%FV) $ENDMA
$RETURN
$C
$C
```

```
$C
$SUBFI AEXP  !  EXPONENTIAL DISTRIBUTION ( WITH 1 PARAMETER )
$MACRO OWN0 $CALCU %Z =%CU(YV)
                 : %FV=%Z/%NU                    $ENDMA
$MACRO OWN3 $CALCU %VA=%FV*%FV                   $ENDMA
$MACRO OWN4 $CALCU %DI=%YV/%FV
                 : %DI=2*(%DI-1-%LOG(%DI)) $ENDMA
$RETURN
$C

$C
$SUBFI ANEG  !  NEGATIVE BINOMIAL DISTRIBUTION
$MACRO OWN1 $CALCU %LP=%LOG(%FV/(%FV+BD))    $ENDMA
$MACRO OWN1 $CALCU %FV= BD/(%EXP(-%LP)-1)    $ENDMA
$MACRO OWN2 $CALCU %DR= BD/(%FV*(%FV+BD))    $ENDMA
$C
$MACRO OWN0 $CALCU %FV=YV+.5                                    $ENDMA
$MACRO OWN3 $CALCU %VA=%FV*(%FV+BD)/BD                          $ENDMA
$MACRO OWN4 $CALCU %DI=2*(%YV*%LOG(%YV/%FV)-
                    (%YV+BD)*%LOG((%YV+BD)/(%FV+BD))) $ENDMA
$RETURN
$C

$C
$SUBFI AGEO  !  GEOMETRIC DISTRIBUTION
$MACRO OWN1 $CALCU %LP=%LOG(%FV/(%FV+1)) $ENDMA
$MACRO OWN1 $CALCU %FV=1/(%EXP(-%LP)-1)  $ENDMA
$MACRO OWN2 $CALCU %DR=1/(%FV*(%FV+1))   $ENDMA
$C
$MACRO OWN0 $CALCU %FV=YV+.5                                $ENDMA
$MACRO OWN3 $CALCU %VA=%FV*(%F+1)                           $ENDMA
$MACRO OWN4 $CALCU %DI=2*(%YV*%LOG(%YV/%FV)-
                    (%YV+1)*%LOG((%YV+1)/(%FV+1))) $ENDMA
$RETURN
$C

$C
$SUBFI ANOI  !  INVERSE NORMAL DISTRIBUTION

$MACRO OWN1 $CALCU %LP=-1/(2*%FV*%FV)    $ENDMA
$MACRO OWN1 $CALCU %FV=%SQRT(-1/(2*%LP)) $ENDMA
$MACRO OWN2 $CALCU %DR=1/(%FV*%FV*%FV)   $ENDMA
$C
$MACRO OWN0 $CALCU %FV=YV+.5             $ENDMA
$MACRO OWN3 $CALCU %Z =%CU(%YV-%FV)**2
                 : %Z =%Z/(%NU-%PL)
                 : %VA=%Z/%DR            $ENDMA
$MACRO OWN4 $CALCU %DI=(%YV-%FV)**2/
                    (%YV*%FV*%FV)   $ENDMA
$FINISH
```

PRELIMINARY DRAFT

Manfred Kuechler

University of Frankfurt
and
Center for Survey Research &
Methodology (ZUMA), Mannheim
West Germany

Jeffrey W. Wides

Southwestern Bell Telephone Company,
St. Louis, Mo.
USA

ECONOMIC PERCEPTIONS AND THE '76 AND '80 PRESIDENTIAL VOTES+

mit einem EXKURS über die programmtechnische Umsetzung der diskutierten Analyse

Paper presented at the 1981 annual meetings of the American Sociological Association in Toronto, Canada.

Session on "Advances in Discrete Multivariate Analysis"
(Organizer: Robert M. Hauser)

Any obscurities of language or infelicities in style are the responsibility of the first-named author.

Introduction

It seems self-evident that there would be a relationship between economic factors and electoral choice. However we define politics, the phenomenon involves conflict or cooperation in the actual or ideal process of distributing public resources. As income, jobs, and living conditions are important public resources, political decisions almost by definition would include consideration of those resources. As one's voting choice is a political decision, it too should probably reflect consideration of economic factors. Yet a relationship between economic factors and vote has not been well established. The conventional wisdoms that an administration's economic performance will significantly effect its electoral performance, and that economic fluctuations are impressively associated with vote fluctuations, have not received consistent support.[1)]

Following Wahlke's (1979:13) recommendation we concentrate on the "micro" level of analysis. By use of survey data we try to relate individual voting decision to variables like voter's personal economic situation, perceived economic competence of candidates etc. always controlling for partisanship. Whereas such individual level analysis seems to be rare in the early seventies, there are some more recent articles most pertinent to our approach (among others Brody and Sniderman 1977, Fiorina 1978, Kinder and Kiewiet 1979, Miller 1978). However, by employing a data analytic technique developed by Grizzle, Starmer and Koch (1969) we claim to analyze our data at the same time statistically sound and meaningful in substantive interpretation.

Summary of '76 findings[2)]

Among a set of "economic variables" candidate's perceived economic competence (in handling inflation and unemployment) proved to be the most influential factor. Perceived competency even outweighed partisanship in direct impact, though in turn partisanship heavily effects competency evaluations (operationalized as a relative measure between candidates). Voter's outlook on his/her own economic situation did not show any significant effect when controlling for other factors found to be relevant. A minor impact was assigned to voter's evaluation of government's economic policy.

While the non-significance of personal outlook confirms the findings of Brody and Sniderman (1977) as well as Kinder and Kiewiet (1979), candidates' economic competence has not been used in a systematic way by other researchers. To the contrary, Miller (1978) dismisses attempts to introduce party economic competency for reasons of multicollinearity. Instead, by use of metric path analytic techniques, he reinstates party identification as the dominant explanatory factor.

Given the relative novelty of our choice of indicators, we feel that its usefulness could best be evaluated by an attempt to replicate our model with the 1980 data.

Data

The data used for the 1980 analysis were taken from the pre/post election panel (C3/C3po) of CPS's election study.[3)] Both candidates

were rated on a four point scale whether they "would solve our economic problems" (V428/V446).[4] A trichotomy was derived by comparing Reagan with Carter ratings, including all missing data cases in the ("indifferent") medium category. For party identification the composite measure offered by CPS was trichotomized counting "independent leaners" as independents.[5] Unfortunately, the question on government's economic policy, that we used in 1976, was not included in the 1980 survey. As a surrogate a more general question on the performance of the Federal government was used (V761). The 9-point rating scale was dichotomized with dissatisfied respondents contrasted with the rest.[6] Finally, as in 1976, several variables on the personal economic situation were explicitly taken into account (V150 ff.).

Methodology

Path analysis is generally an adequate tool for examining a complex structure among a set of interval-level variables. If the dependent variable is not metric, however, basic assumptions for doing path analysis are violated (see e.g. Duncan, 1975, pp. 160-161); neither normality in the distribution of the dependent variable nor homoscedasticity of error terms can be rightfully assumed. A number of approaches to the causal analyses of nonmetric data have been put forward during the last decade, most notably in the field of social sciences by Leo A. Goodman (e.g. 1978), and textbooks for use in practical research are available (among others Reynolds, 1977; Fienberg, 1977; Kuechler, 1979; Forthofer and Lehnen 1981).

Virtually all of the nonmetric techniques can be understood as modifications of ordinary least squares regression analysis. They divide the sample into subpopulations that are homogeneous in respect to the values of the independent variables and then examine the distribution of the dependent variable within each subpopulation. One works on the aggregate level with subpopulations as cases and uses a (modified) regression equation to specify differences among the independent variable(s). The approach is very similar to analysis of variance, or, more simply, to the investigation of percentage differences in a four-fold table. Since the dependent variable is nonmetric, group differences are described through a more general "metrization" of the dependent variable rather than group means.

Let us be specific and consider the simple case of a dichotomous dependent variable, (e.g. Presidential vote) as in Table 1. Here each conditional distribution can be fully described by a single number, the proportion of cases falling into a fixed category (denoted by p). It can also be described by the so called log-odds i.e. the natural logarithm of the ratio of the proportions falling into the first and second category respectively. This second approach is known as log-linear modelling and advocated by Goodman, among others.[7] Its advantage is that the range of metrization is not limited, hence any estimates can never exceed the range of substantive meaning. At the same time, the formal and mathematical nature of the "effects" obtained by this method make it extremely difficult to disseminate the results to an audience outside the world of methodological specialists.

TABLE 1:

FREQUENCY DISTRIBUTION: VOTING DECISION BY (PARTY-ID X ECON.COMP.X GOV.POL.)

				VOTING DECISION			
	PARTY-ID	COMPETENCE	GOV.POL.	REAGAN	CARTER	TOTAL	% REAGAN
1	DEM	REAGAN	POOR	17	4	21	81.0
2	DEM	REAGAN	FAIR	35	12	47	74.5
3	DEM	INDIFF	POOR	15	24	39	38.5
4	DEM	INDIFF	FAIR	12	98	110	10.9
5	DEM	CARTER	POOR	3	27	30	10.0
6	DEM	CARTER	FAIR	4	110	114	3.5
7	IND	REAGAN	POOR	66	6	72	91.7
8	IND	REAGAN	FAIR	58	9	67	86.6
9	IND	INDIFF	POOR	14	11	25	56.0
10	IND	INDIFF	FAIR	26	28	54	48.1
11	IND	CARTER	POOR	2	10	12	16.7
12	IND	CARTER	FAIR	5	16	21	23.8
13	REP	REAGAN	POOR	93	0(.5)	93	99.5
14	REP	REAGAN	FAIR	83	2	85	97.6
15	REP	INDIFF	POOR	28	1	29	96.6
16	REP	INDIFF	FAIR	24	3	27	88.9
17	REP	CARTER	POOR	1	1	2	50.0
18	REP	CARTER	FAIR	2	5	7	28.6
				488	367	855	

PEARSON CHI-SQUARE: 509.77 DF = 17
CRAMER'S V = 0.772

DATA SOURCE: CPS-ELECTION STUDY 1980 (C3/C3PO)

We therefore have decided to follow the approach developed by Grizzle, Starmer, and Koch (1969), which offers a rather general frame to include various different ways of analyzing one's data (linear and log-linear models among many others).

The starting point for a GSK analysis is an SxR frequency table, where S is the number of subpopulations, i.e. the product of numbers of categories for all independent variables involved, and R is the number of categories for the dependent variable. In our case three independent variables (two trichotomies, one dichotomy) and a dichotomous dependent variable would produce a 18x2 table. The GSK approach is not primarily concerned with measuring the overall impact of the set of independent variables on the dependent one, but rather assumes that the initial frequency table is well worth further investigation. We measure the overall impact of the independent variables by using the familiar association measure suggested by Cramer and usually denoted by V. With a dichotomy as dependent variable it can be shown that the coefficient of determination obtained by ordinary regression in a saturated model is numerically identical to V^2.[8)]

Before proceeding with the analysis of the data at hand, we would like to briefly outline the GSK approach in more formal terms. Readers with less interest in statistical details might want to skip this paragraph.[9)]

In the GSK approach the subpopulations are assumed to be independent samples following a multinomial distribution. Each subpopulation i is characterized by a vector $p_i = (p_{i1}, \ldots p_{ir})$ containing the proportions of cases falling into each of the response categories. Likewise P is a column vector of order rs containing all proportions ordered by rows first.

In general, F(P) is considered as dependent variable, where F can be any of a rather large class of functions. In the simplest case F is a linear function, which in vector notation can be written

$$F(P) = A \cdot P$$

In our case A would be of blockdiagonal form with (1 0) as diagonal element, selecting the first proportion (percentage Reagan) for each subpopulation. The log-odds model could be defined as

$$F(P) = K(\ln A \cdot P)$$

with A the identity matrix and K blockdiagonal with (1 -1). Function can be as complex as

$$F(P) = J \exp (L \ln(Q \exp K (\ln A \cdot P))),$$

e.g. to model association coefficients of subtables as functions.[10)]

By means of the delta method the corresponding variance-covariance matrix V = V(F) can be estimated.

Whatever the function is, the user will then specify a set of dummy variables (defined on the subpopulations) to model the effects

of the independent variables. These dummy variables are combined into the "design matrix" X to solve a general linear model:

$$F = F(P) = X \cdot b$$

Again, this equation is to be read as a matrix equation.

To account for heteroscedasticity a weighted least square technique is employed to render

$$b = (X'V^{-1}X)^{-1} \quad X'V^{-1}F \ .$$

The goodness-of-fit of a particular model can be evaluated by looking at the Wald statistic

$$W = F'V^{-1}F - b'(X'V^{-1}X)^{-1}b,$$

which is approximately distributed as chi-square with df = t-v and t the order of F, v the rank of X.
Also, any linear hypothethis

$$Cb = 0$$

can be tested by looking at

$$H = (Cb)' \ (CX'V^{-1}XC')^{-1}(Cb),$$

which again is approximately distributed as chi-square with as many degrees of freedom as is the rank of C. Particularly, the effect of each single effect can be evaluated.[11)]

Though the statistical foundations as well as the very general formulation of this approach may look frightening to researchers more interested in applied research than formal statistics, the interpretation of the models arrived at is rather appealing to intuitive thinking.

EXKURS: NONMET - Eingabe und Ausgabe
für die diskutierte Analyse (Teil A)

Eingabe

1. /DO $ZUMA.NONMET,(SPACE=35000)
anlagenspezifischer Aufruf; der benötigte Speicherplatz hängt von der Ausgangskreuztabelle und der Zahl der daraus abgeleiteten "Funktionen" ab.
Für SIEMENS-Anlagen kann der default-Wert durch die zusätzliche Angabe
.....,(SPACE=nnn)
verändert werden.

2. TITLE CPS C3/C3PO DATA
Die Angabe TITLE in Spalten 1-5 ist zwingend vorgeschrieben (im Batch-Betrieb). Der Rest der Karte enthält einen beliebigen Text.

3. NR=2,NP=18,OUT=3/
NR und NP definieren das Format der Ausgangstabelle; OUT=3 bewirkt eine verkürzte Ausgabe (ohne Kovarianzmatrizen).

4a. 17 4 35 .....
b.
c.
Freiformatige zeilenweise Eingabe der Häufigkeiten der Ausgangstabelle.

5. LE=P(3),C(3),G(2)/
Identifikation der Einflußfaktoren in der Reihenfolge der Tabelle, die Zahl in Klammern gibt die Zahl der Ausprägungen pro Faktor an.
Als Default wird das MAIN-Modell gerechnet, das nur die Haupteffekte umfaßt.

```
6.   SINGLE
```

Anforderung von Tests für jeden einzelnen Parameter; statt 'SINGLE' kann auch '990' angegeben werden.

```
7.   REAN
```

Der Vektor F(P) soll mit neuer Design-Matrix reanalysiert werden.

```
8.   MAIN=P1,C1,G;NACT=P1C/

9.   SINGLE

10.  REAN

11.  MAIN=C1,P1<C2,P1<C13,G<C2/          "Bestes" Modell

12.  SINGLE

13.  TITLE  PARALLEL  AUSWERTUNG  LOGIT......

14.  NR=2,NP=18,A=LOGIT,OUT=3/
```

Durch "A=LOGIT" werden die log-odds als zu analysierende Funktionen bestimmt (anstelle der Anteilswerte).

```
15a.      (wie 4a,b,c,.....)
  b.
  c.

16.  LE=P(3),C(3),G(2);BA=P1,C1,G1,EF=CO/
```

Durch 'EF=CO' werden sogenannte 'cornered Effekts' (Dummy-Variablen mit 1/0-Kodierung) betrachtet.
Als Default wird bei NONMET stets die letzte Ausprägung nicht explizit durch eine Dummy-Variable repräsentiert. Bei GLIM ist es gerade die erste.
Durch die 'BA=.....'Angabe kann die Basiskategorie jedoch frei gewählt werden. Hier wird die Analogie zum GLIM-Default hergestellt.

```
17.  SINGLE

18.  END
```

```
      ****  NONMET  ****

VERSION:  6.11 (APR. 15, 1981)

TODAY'S DATE: 19/11/81

  CPS C3/C3PO DATA KUEJ11A OUTPUT

NRESP (R) =  2

NPOP  (S) = 18

MATRICES USED:
     A* (U*=  1, U = 18)
     X
     C

PROBLEM USES  13184 OF THE  35000 BYTES AVAILABLE

FREQUENCY TABLE (S X R)
            1          2
  1        17.         4.
  2        35.        12.
  3        15.        24.
  4        12.        98.
  5         3.        27.
  6         4.       110.
  7        66.         6.
  8        58.         9.
  9        14.        11.
 10        26.        28.
 11         2.        10.
 12         5.        16.
 13        93.         0.
 14        83.         2.
 15        28.         1.
 16        24.         3.
 17         1.         1.
 18         2.         5.

PROBABILITY TABLE (S X R)
            1          2
  1    0.80952    0.19048
  2    0.74468    0.25532
  3    0.38462    0.61538
  4    0.10909    0.89091
  5    0.10000    0.90000
  6    0.03509    0.96491
  7    0.91667    0.08333
  8    0.86567    0.13433
  9    0.56000    0.44000
 10    0.48148    0.51852
 11    0.16667    0.83333
 12    0.23810    0.76190
 13    0.99465    0.00535
 14    0.97647    0.02353
 15    0.96552    0.03448
 16    0.88889    0.11111
 17    0.50000    0.50000
 18    0.28571    0.71429

A* MATRIX IS DEFAULT RESPONSE FUNCTION
```

Printback der Eingabe-Daten

Es zeigt sich, daß der default-Wert von 20 000 ausgereicht hätte.

Printback der Eingabe-Daten

zeilenweise Anteilsberechberechnung (vgl. letzte Spalte von TABLE 1)

Default:
Die (r-1) ersten Anteilswerte gelten als 'Funktionen', die analysiert werden.

```
GENERATE X MATRIX:
   LE=P(3),C(3),G(2)/
```

DESIGN MATRIX (T X V)

	1	2	3	4	5	6
	MEAN	P1	P2	C1	C2	G
1	1.00	1.00	0.00	1.00	0.00	1.00
2	1.00	1.00	0.00	1.00	0.00	-1.00
3	1.00	1.00	0.00	0.00	1.00	1.00
4	1.00	1.00	0.00	0.00	1.00	-1.00
5	1.00	1.00	0.00	-1.00	-1.00	1.00
6	1.00	1.00	0.00	-1.00	-1.00	-1.00
7	1.00	0.00	1.00	1.00	0.00	1.00
8	1.00	0.00	1.00	1.00	0.00	-1.00
9	1.00	0.00	1.00	0.00	1.00	1.00
10	1.00	0.00	1.00	0.00	1.00	-1.00
11	1.00	0.00	1.00	-1.00	-1.00	1.00
12	1.00	0.00	1.00	-1.00	-1.00	-1.00
13	1.00	-1.00	-1.00	1.00	0.00	1.00
14	1.00	-1.00	-1.00	1.00	0.00	-1.00
15	1.00	-1.00	-1.00	0.00	1.00	1.00
16	1.00	-1.00	-1.00	0.00	1.00	-1.00
17	1.00	-1.00	-1.00	-1.00	-1.00	1.00
18	1.00	-1.00	-1.00	-1.00	-1.00	-1.00

```
CHI-SQUARE DUE TO ERROR =    108.9646 DF =     12      P = -.0000
```

NONMET-Default der Design-Matrix: Zentrierte Effekte (1/-1 Kodierung; jeweils letzte Ausprägung wird nicht explizit repräsentiert.
Die WALD-Statistik (gewichtete Quadratsumme der Residuen) - im Ausdruck CHI-SQUARE DUE TO ERROR - ist hochsignifikant, d.h. das Modell beschreibt die empirische Konstellation nur unzureichend.

```
USER REQUESTED CONTRASTS:
```

```
TESTS OF INDIVIDUAL PARAMETERS:
```

PARAMETER	B	VARIANCE	CHI SQUARE	P
1 MEAN	0.5535D+00	0.1152D-03	2658.63	0.00000
2 P1	-0.2801D+00	0.3737D-03	209.99	0.00000
3 P2	0.6912D-01	0.2694D-03	17.74	0.00003
4 C1	0.2209D+00	0.2641D-03	184.69	0.00000
5 C2	0.5597D-02	0.1950D-03	0.16	0.68852
6 G	0.2242D-01	0.5806D-04	8.65	0.00326

Die Effekte (Spalte P) sind signifikant von Null verschieden (Ausnahme Effekt Nr. 5 gleich C2).

RE ANALYSIS:

GENERATE X MATRIX:
MAIN=P1,C1,G;NACT=P1C/

DESIGN MATRIX (T X V)

	1 MEAN	2 P1	3 C1	4 G	5 P1C1	6 P1C2
1	1.00	1.00	1.00	1.00	1.00	0.00
2	1.00	1.00	1.00	-1.00	1.00	0.00
3	1.00	1.00	0.00	1.00	0.00	1.00
4	1.00	1.00	0.00	-1.00	0.00	1.00
5	1.00	1.00	-1.00	1.00	-1.00	-1.00
6	1.00	1.00	-1.00	-1.00	-1.00	-1.00
7	1.00	0.00	1.00	1.00	0.00	0.00
8	1.00	0.00	1.00	-1.00	0.00	0.00
9	1.00	0.00	0.00	1.00	0.00	0.00
10	1.00	0.00	0.00	-1.00	0.00	0.00
11	1.00	0.00	-1.00	1.00	0.00	0.00
12	1.00	0.00	-1.00	-1.00	0.00	0.00
13	1.00	-1.00	1.00	1.00	-1.00	0.00
14	1.00	-1.00	1.00	-1.00	-1.00	0.00
15	1.00	-1.00	0.00	1.00	0.00	-1.00
16	1.00	-1.00	0.00	-1.00	0.00	-1.00
17	1.00	-1.00	-1.00	1.00	1.00	1.00
18	1.00	-1.00	-1.00	-1.00	1.00	1.00

CHI-SQUARE DUE TO ERROR = 11.4304 DF = 12 P = 0.4924

Die Anpassung des um die Intraktionswirkung zwischen Partei-Identifikation und Kompetenzbewertung erweiterten Modells ist bereits ausgezeichnet.

TESTS OF INDIVIDUAL PARAMETERS:

PARAMETER	B	VARIANCE	CHI SQUARE	P
1 MEAN	0.5440D+00	0.2469D-03	1198.70	0.00000
2 P1	-0.2106D+00	0.2866D-03	154.72	0.00000
3 C1	0.3405D+00	0.4515D-03	256.75	0.00000
4 G	0.1868D-01	0.5821D-04	6.00	0.01434
5 P1C1	0.1154D+00	0.2606D-03	51.12	-0.00000
6 P1C2	-0.1789D+00	0.3958D-03	80.85	-0.00000

Alle Effekte sind signifikant, aber

- Effekt Nr. 4 (G) ist substantiell recht gering: 0.0186, also durchgängig nur etwa 3,5 Prozentpunkte Unterschied je nach Bewertung der Regierung (G).
- Effekte Nr. 5 und 6 lassen sich so nur schwer substantiell interpretieren.

REANALYSIS:

```
GENERATE X MATRIX:
   MAIN=C1,P1<C2,P1<C13,G<C2/

DESIGN MATRIX (T X V)
```

	1 MEAN	2 C1	3 P1<C2	4 P1<C13	5 G<C2
1	1.00	1.00	0.00	1.00	0.00
2	1.00	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00	1.00
4	1.00	0.00	1.00	0.00	-1.00
5	1.00	-1.00	0.00	1.00	0.00
6	1.00	-1.00	0.00	1.00	0.00
7	1.00	1.00	0.00	0.00	0.00
8	1.00	1.00	0.00	0.00	0.00
9	1.00	0.00	0.00	0.00	1.00
10	1.00	0.00	0.00	0.00	-1.00
11	1.00	-1.00	0.00	0.00	0.00
12	1.00	-1.00	0.00	0.00	0.00
13	1.00	1.00	0.00	-1.00	0.00
14	1.00	1.00	0.00	-1.00	0.00
15	1.00	0.00	-1.00	0.00	1.00
16	1.00	0.00	-1.00	0.00	-1.00
17	1.00	-1.00	0.00	-1.00	0.00
18	1.00	-1.00	0.00	-1.00	0.00

```
CHI-SQUARE DUE TO ERROR =     11.5445 DF =     13      P = 0.5653
```

Ein einfacheres Modell (1 Effekt = 1 Freiheitsgrad weniger) erbringt fast den gleichen CHI-Quadrat Wert, ist also relativ gesehen besser angepaßt.
Zur Modellierung der Intraktionswirkungen wurden nur konditionale Effekte benutzt. Der Partei-Identifikationseffekt wird also getrennt für in der Kompetenzbewerung Ambivalente (C2) und Entschiedene (C1 und C3 = C13) berechnet.
Der Regierungspolitik-Effekt (G) wird deutlicher, wenn man ihn auf Kompetenz-Ambivalente beschränkt:

```
TESTS OF INDIVIDUAL PARAMETERS:
```

PARAMETER	B	VARIANCE	CHI SQUARE	P
1 MEAN	0.5239D+00	0.6196D-04	4430.31	0.00000
2 C1	0.3607D+00	0.3302D-03	393.97	0.00000
3 P1<C2	-0.3571D+00	0.6441D-03	197.98	0.00000
4 P1<C13	-0.1094D+00	0.3147D-03	38.01	-0.00000
5 G<C2	0.6725D-01	0.5726D-03	7.90	0.00494

```
SKIPPING TO NEXT PROBLEM
```

Das letzte Modell kann zumindest formal als ein 'bestes Modell' betrachtet werden. Es ist einfach (nur 5 von 18 Freiheitsgraden), es ist den empirischen Daten hervorragend angepaßt ($P \geqslant 0.50$), alle berechneten Effekte sind statistisch signifikant und von der Größenordnung her substantiell relevant.
Aus den Effekten können nun auch Predictorwerte für die einzelnen Subpopulationen berechnet werden, die das NONMET-Programm automatisch mit ausgibt:

F(P) PREDICTED FROM FITTED MODEL

1	2	3	4	5
0.77526D+00	0.77526D+00	0.23410D+00	0.99588D-01	0.53876D-01
9	10	11	12	13
0.59118D+00	0.45668D+00	0.16324D+00	0.16324D+00	0.99398D+00
17	18			
0.27260D+00	0.27260D+00			

6	7	8
0.53876D-01	0.88462D+00	0.88462D+00
14	15	16
0.99398D+00	0.94827D+00	0.81376D+00

Alle Predictorwerte liegen im zulässigen Bereich zwischen 0 und 1, einige allerdings sehr dicht an diesen Grenzen (z.B. Subpopulation Nr. 4, 13, 14). Dies war zu erwarten, da die Ausgangstabelle bereits sehr schief verteilt ist (vgl. PROBABILITY TABLE).
Nähme man zu den Predictorwerten noch die Konfidenzintervalle hinzu ($\pm 2\sqrt{\text{Var } F(P)}$; für i = 1,.....,18), die man aus der Kovarianzmatrix für F(P) leicht berechnen kann ('OUT = 3' bei der Eingabe fortlassen), so würden diese Intervalle zum Teil in den nicht sinnvoll definierten Bereich hineinreichen.

Aus diesem Grund bevorzugen eine Reihe von Statistikern log-lineare Modelle, bei denen derartige Inkonsistenzen per Konstruktion ausgeschlossen werden. Auch in GLIM sind standardmäßig für dichotome abhängige Variable log-odds (oder logits) vorgesehen. In ARMINGERs Papier wird jedoch demonstriert, daß auch eine lineare NONMET-Analyse mit dem GLIM-Programm gerechnet werden kann und zu identischen Ergebnissen führt.
Hier wird umgekehrt die GLIM-Standard-Option innerhalb von NONMET gerechnet, wobei das Schätzverfahren - wie bei ARMINGER dargestellt - differiert; bei NONMET also nur der erste Schritt eines Iterationsverfahrens gerechnet wird:

```
A* MATRIX IS DEFAULT LOGIT MATRIX

K MATRIX IS DEFAULT BLOCK DIAGONAL LOGIT MATRIX

F(P)        LOG MODEL
          1             2             3             4             5
 0.14469D+01   0.10704D+01  -0.47000D+00  -0.21001D+01  -0.21972D+01
          9            10            11            12            13
 0.24116D+00  -0.74108D-01  -0.16094D+01  -0.11632D+01   0.52258D+01
         17            18
 0.00000D+00  -0.91629D+00

          6             7             8
-0.33142D+01   0.23979D+01   0.18632D+01
         14            15            16
 0.37257D+01   0.33322D+01   0.20794D+01
```

```
GENERATE X MATRIX:
   LE=P(3),C(3),G(2);BA=P1,C1,G1;EF=CO/
```

DESIGN MATRIX (T X V)

	1	2	3	4	5	6
	MEAN	P1	P2	C1	C2	G
1	1.00	0.00	0.00	0.00	0.00	0.00
2	1.00	0.00	0.00	0.00	0.00	1.00
3	1.00	0.00	0.00	1.00	0.00	0.00
4	1.00	0.00	0.00	1.00	0.00	1.00
5	1.00	0.00	0.00	0.00	1.00	0.00
6	1.00	0.00	0.00	0.00	1.00	1.00
7	1.00	1.00	0.00	0.00	0.00	0.00
8	1.00	1.00	0.00	0.00	0.00	1.00
9	1.00	1.00	0.00	1.00	0.00	0.00
10	1.00	1.00	0.00	1.00	0.00	1.00
11	1.00	1.00	0.00	0.00	1.00	0.00
12	1.00	1.00	0.00	0.00	1.00	1.00
13	1.00	0.00	1.00	0.00	0.00	0.00
14	1.00	0.00	1.00	0.00	0.00	1.00
15	1.00	0.00	1.00	1.00	0.00	0.00
16	1.00	0.00	1.00	1.00	0.00	1.00
17	1.00	0.00	1.00	0.00	1.00	0.00
18	1.00	0.00	1.00	0.00	1.00	1.00

CHI-SQUARE DUE TO ERROR = 12.7591 DF = 12 P = 0.3868

TESTS OF INDIVIDUAL PARAMETERS:

PARAMETER	B	VARIANCE	CHI SQUARE	P
1 MEAN	0.1466D+01	0.7175D-01	29.97	0.00000
2 P1	0.1278D+01	0.5210D-01	31.33	0.00000
3 P2	0.3217D+01	0.1523D+00	67.97	-0.00000
4 C1	-0.2259D+01	0.5911D-01	86.35	-0.00000
5 C2	-0.3869D+01	0.1080D+00	138.58	0.00000
6 G	-0.7477D+00	0.5164D-01	10.82	0.00100

Log-linear modelliert ist bereits das Haupteffekt-Modell gut angepaßt. Es läßt sich auch nicht weiter vereinfachen, da alle Effekte hochgradig signifikant sind. Ebenso wie beim linearen Umsatz können Predictorwerte (für log-odds) berechnet, die mit Hilfe der Transformation $e^X/(1+e^X)$ in Anteilswerte rückverwandelt werden können. Liegt das Forschungsinteresse hauptsächlich bei den Predictorwerten (also weniger bei den Effekten) fällt die mindere Intuitivität der log-odds also kaum ins Gewicht, da sie dann nur in Zwischenrechnungen auftauchen.

```
$C  KUECHLERS DATEN MIT GLIM GERECHNET LOGIT MODELL
$ UNITS 18 $FAC P 3 C 3 G 2
$ DATA P C G R N
$ READ
1 1  1 17 21
1 1  2 35 47
1 2  1 15 39
1 2  2 12 110
1 3  1 3 30
 1 3 2 4 114
2 1  1 66 72
2 1  2 58 67
2 2  1 14 25
2 2  2 26 54
2 3  1 2 12
2 3  2 5 21
3 1  1 93 93
3 1  2 83 85
3 2  1 28 29
3 2  2 24 27
3 3  1 1 2
3 3  2 2 7
$YVAR R $ERR B N $LINK G
$FIT P+C+G $DIS MERVS
         SCALED
CYCLE    DEVIANCE      DF
   4      14.67        12

Y-VARIATE R
ERROR BINOMIAL LINK LOGIT
BINOMIAL DENOMINATOR N

LINEAR PREDICTOR
% GM P C G

       ESTIMATE      S.E.      PARAMETER
     1   1.540     0.2774      %GM
     2   1.296     0.2290      P(2)
     3   3.368     0.3593      P(3)
     4  -2.329     0.2435      C(2)
     5  -3.999     0.3436      C(3)
     6 -0.8172     0.2292      G(2)
    SCALE PARAMETER TAKEN AS  1.000

    UNIT OBSERVED OUT OF  FITTED   RESIDUAL
     1       17      21   17.29   -0.1668
     2       35      47   31.64    1.046
     3       15      39   12.18    0.9742
     4       12     110   18.38   -1.630
     5        3      30   2.363    0.4319
     6        4     114   4.148   -0.7412E-01
     7       66      72   68.01    -1.035
     8       58      67   59.14   -0.4339
     9       14      25   15.60   -0.6612
    10       26      54   22.84    0.8696
    11        2      12   2.857   -0.5808
    12        5      21   2.547    1.640
    13       93      93   92.32    0.8291
    14       83      85   83.60   -0.5124
    15       28      29   26.95    0.7582
    16       24      27   23.04    0.5215
    17        1       2   1.425   -0.6648
    18        2       7   3.660    -1.256
```

Zum Vergleich:

Die Ausgabe der gleichen Analyse mit GLIM gerechnet. Die mit $ beginnenden Karten wie die Datenkarten bilden die Eingabe

Bei den Effekten treten kleinere Abweichungen auf, da hier 4 Zyklen des Iterationsprozess durchlaufen worden sind.

Eine benutzerfreundlichere Ausgabe (z.B. Signifikanzen für die Effekte) kann durch eigene Programmierergänzungen (MACROS) erzielt werden.

Results

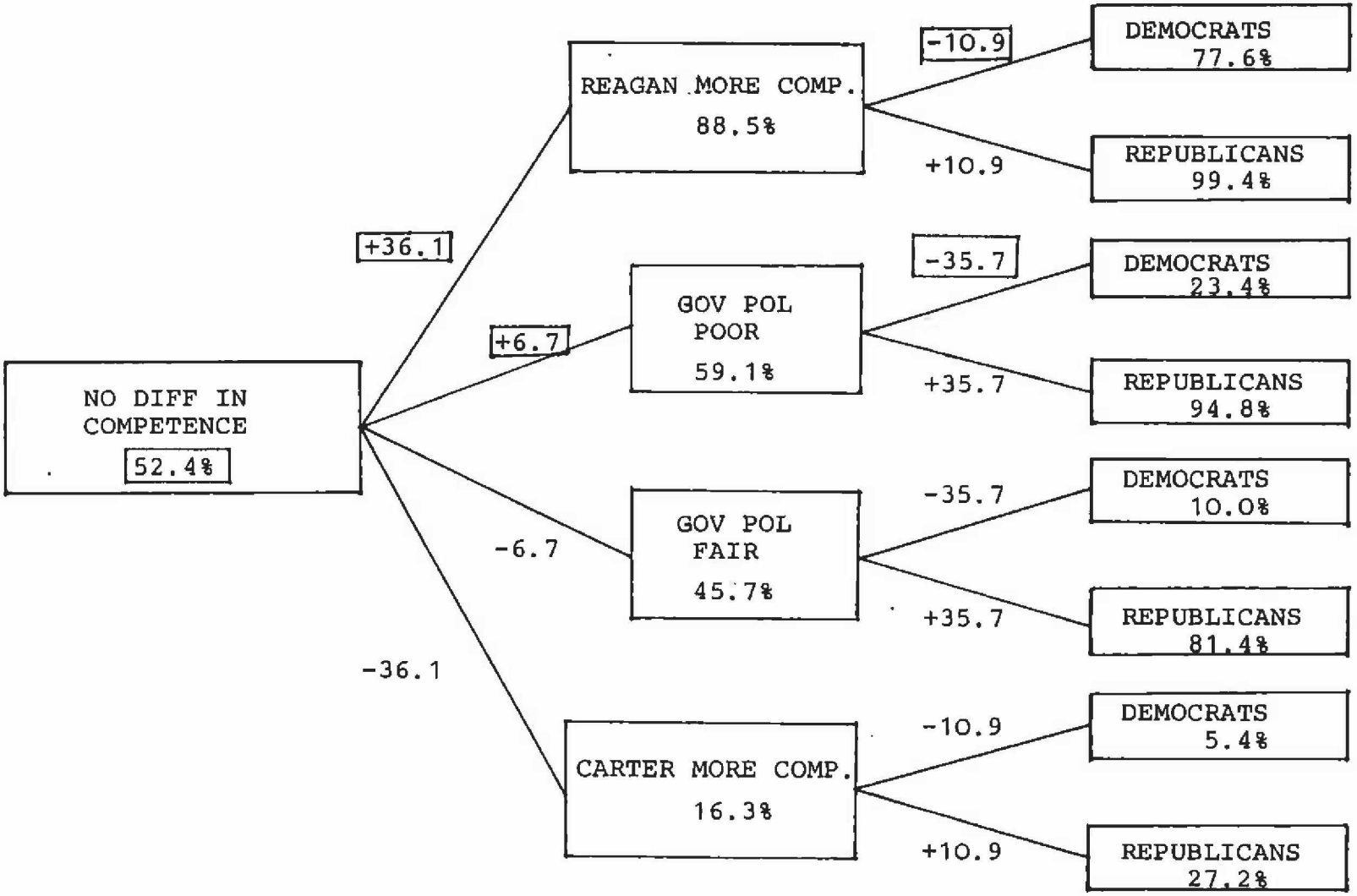
The results of our analysis of Table 1 are graphically displayed in Table 2. With F(P) of order 18 (i.e. one proportion for each of the 18 subpopulations) we were able to achieve excellent fit (P above .5) with just 5 parameters, i.e. 13 degrees of freedom. Obviously, candidates' competence is the predominant factor. Other things being equal, the percentage of voting for Reagan in subpopulations where Reagan is considered more competent is 72 (= 2x36.1) percentage points higher than in subgroups where Carter is seen as more competent. The impact of party identification is contingent upon competence evaluation. Only when a difference between candidates' competency is not perceived is party identification as strong as competence.

Also, evaluation of government policy has limited impact and is confined to "indifferent" subpopulations. In addition, the possible impact of "personal situation" variables was investigated by looking at a summated chi-square, crosstabulating the perspective variables with the voting decision and controlling for the combined impact of the three independent variables of Table 1.[12)]

Hence, we were able to replicate our model of the '76 analysis with the '80 data. A more detailed comparison is given in Table 3. However to fully understand the notation used it might help to look at Table 4 first, where the "design matrix" used is displayed in detail. As a matter of taste - not statistical necessity - we prefer the effect coding, i.e. assigning '1' to the category in question, '-1' to a selected 'base category' and '0' elsewhere.[13)]

TABLE 2:

GRAPHICAL PRESENTATION OF BEST MODEL FOR RERCENTAGE REAGAN VOTE '80



$\chi^2 = 11.54$

DF = 13 (out of 18)

P = .5653

TABLE 3: COMPARISON OF '76 AND '80 RESULTS

Effects	1976	1980	
MEAN	54.7%	52.4%	
P1 < C13	-10.2%	-10.9%	
P1 < C2	-25.8%	-35.7%	
C1	33.3%	36.1%	(C1)
C3	-39.3%		
G < C3	2.3%	6.7%	(G < C2)
x^2	14.82	11.54	
df	12	13	
P	.2516	.5653	

All effects significant on at least 1%-level, except G < C3 (1976) on 5%-level only.

In both 1976 and 1980 the Republican candidate's voting percentage is analyzed, in order to show the effect of G on the incumbent, the 1976 order of G categories has been reversed for 1980.

The overall pattern appears to be amazingly stable, though partisanship seems gaining in power again after the exceptional election of 1972.[14)]

Some remarks on methodology

Before proceeding to a more refined "path analytic" approach to the data constellation at hand, we would like to comment on the methodology employed in respect to substantive applications.

Firstly, this kind of analysis is an analysis on types rather than an attempt to reconstruct the marginal distribution of votes. The number of cases in each subpopulation remains largely ignored.[15)] As a consequence any multicollinearity problem present is resolved by the very approach of transferring the level of analysis to types (subpopulations), provided sufficient subpopulation sizes (see below).

Secondly, there is no automatic way in arriving at a "best model", i.e. a model with sufficient fit ($p > .25$ as a rule of thumb), all effects significant and no significant effect outside the model. A design matrix as shown in Table 4 is usually arrived at through a search process guided by substantive insight as well as statistical criteria.

Thirdly, interaction of independent variables in their impact on the dependent variable, which is quite frequent in contingency table analysis, can be modelled in many different ways. Conditional effects

TABLE 4: DESIGN-MATRIX USED IN '80 ANALYSIS

	PARTY	COMP	GOV	MEAN	P1 < C13
1	DEM	RGN	POOR	1	1
2	DEM	RGN	FAIR	1	1
3	DEM	IND	POOR	1	0
4	DEM	IND	FAIR	1	0
5	DEM	CTR	POOR	1	1
6	DEM	CTR	FAIR	1	1
7	IND	RGN	POOR	1	0
8	IND	RGN	FAIR	1	0
9	IND	IND	POOR	1	0
10	IND	IND	FAIR	1	0
11	IND	CTR	POOR	1	0
12	IND	CTR	FAIR	1	0
13	REP	RGN	POOR	1	-1
14	REP	RGN	FAIR	1	-1
15	REP	IND	POOR	1	0
16	REP	IND	FAIR	1	0
17	REP	CTR	POOR	1	-1
18	REP	CTR	FAIR	1	-1

P1 < C2	C1	G < C2
0	1	0
0	1	0
1	0	1
1	0	-1
0	-1	0
0	-1	0
0	1	0
0	1	0
0	0	1
0	0	-1
0	-1	0
0	-1	0
0	1	0
0	1	0
-1	0	1
-1	0	-1
0	-1	0
0	-1	0

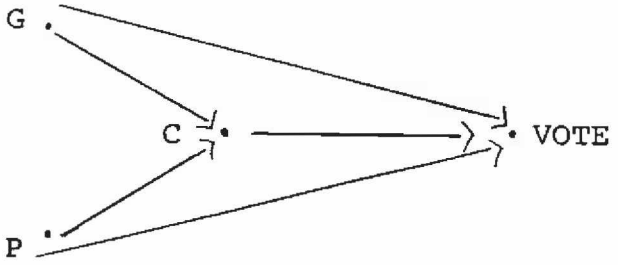
seem to be the most amenable to substantive interpretations. However, it is the researcher's choice to allocate the role of the "condition variable".[16)]

Fourthly, though the GSK approach rests on rather weak assumptions in general, the implicit use of large sample theory requires certain subpopulation sizes. As always, competing rules of thumb are available. Ideally all subpopulations should be of size 20 and above, sizes below 10 should be avoided.[17)] Clearly, in our case this rule is not met. However, we will show below that the results are adequately warranted. These size requirements restrict the number of independent variables and the number of categories per independent variable. Thus grouping of polytomous variables will be necessary in most cases when working with survey data.[18)]

A more complex model

It seems sound substantively to claim that party identification and evaluation of government policy are both antecedent to the perception of candidates' economic competence. This perspective has been neglected so far. All independent variables were treated as on the same level. In our '76 analysis we did a separate analysis on candidates' competency as dependent variable. However, these analyses can be done simultaneously using the GSK approach.[19)]

In our case we postulated the following overall structure:



Due to various interaction effects (modelled as conditional effects) and due to the non-dichotomous character of some of the variables involved (P, C), the relationship between two variables cannot be described by a single number as in metric path analysis.

Also, as Swafford (1980:683) lucidly remarks, a very nice feature of ordinary path analysis is lost when dealing with categorical data: There is no equivalent formula for decomposing a total effect into a direct and the indirect impacts via various path coefficients. Nevertheless, it might be useful to have a test of fit of the overall constellation, i.e. a test of a set of simultaneous equations.

Since we are dealing with a recursive system (P and G are seen as antecedent to C, but not vice versa), C (to be exact: proportions of relative competence) should be estimated by using the multidimensional frequency distribution of (P,G,C) only, while VOTE proportions should be estimated using the four-dimensional count. Using the GSK approach we start out with a 6 x 6 table (G,P as independent; C,V as dependent), as displayed in Table 5. Note, that

TABLE 5:

FREQUENCY DISTRIBUTION: (VOTING DECISION X COMPETENCE) BY (PARTY-ID X GOV.POL.)

		R E A G A N			C A R T E R			(V)	
PARTY-ID	GOVPOL	RGN	IND	CTR	RGN	IND	CTR	(C)	TOTAL
DEM	POOR	17	15	3	4	24	27		90
DEM	FAIR	35	12	4	12	98	110		271
IND	POOR	66	14	2	6	11	10		109
IND	FAIR	58	26	5	9	28	16		142
REP	POOR	93	28	1	0	1	1		124
REP	FAIR	83	24	2	2	3	5		119
		352	119	17	33	165	169		855

the frequencies of Table 1 and Table 5 are identical, only that we have a composite dependent variable in the latter case, allowing for 6 different responses in each of the 6 subpopulations. If p_{ij} denotes the proportion of cases in the i-th subpopulation falling into category j, we can define the following functions on the proportions' vector P (as introduced above):

(1) $f_1 = p_1+p_4$ = proportion of RGN in (P,G,C)

(2) $f_2 = p_3+p_6$ = proportion of CTR in (P,G,C)

(3) $f_3 = p_1/(p_1+p_4)$ = proportion of Reagan in (P,G,C,V) when C = 1

(4) $f_4 = p_2/(p_1+p_5)$ = proportion of Reagan in (P,G,C,V) when C = 2

(5) $f_5 = p_2/(p_3+p_6)$ = proportion of Reagan in (P,G,C,V) when C = 3

Since the same five functions are defined for each subpopulation, the subscript i has been omitted. (P,G,C) denotes the three-dimensional (collapsed) count, (P,G,C,V) the full four-dimensional count.

To fit this into the general notational frame, F(P) can be expressed in matrix terms:

$$F(P) = Q \exp (K \cdot (\ln A \cdot P))$$

with

$$P_{36,1} = \begin{bmatrix} p_{11} \\ \vdots \\ p_{16} \\ \vdots \\ p_{61} \\ \vdots \\ p_{66} \end{bmatrix}$$

$$A_{36,36} = \begin{bmatrix} A^* & & 0 \\ & \ddots & \\ 0 & & A^* \end{bmatrix} \qquad A^*_{6,6} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$K_{30,36} = \begin{bmatrix} K^* & & 0 \\ & \ddots & \\ 0 & & K^* \end{bmatrix} \qquad K^*_{5,6} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

$Q_{30,30}$ = Identity Matrix

A design matrix X is constructed in basically combining the design matrices of the two separate analyses[20] filling in zero for those rows that relate to the perspective "other" functions. Thus by re-arranging the rows X would assume block-diagonal form.

The goodness-of-fit test renders $X^2 = 18.54$ with df = 18 and P = 0.4207. Hence our simultaneous equation model shows excellent fit.

Inspection of the variance-covariance matrix for F(P) shows that

$$\text{cov}\ (f_i, f_j) \approx 0 \qquad i \in \{1,2\} \ , \quad j \in \{3,4,5\}$$

Hence the coefficients estimated should be equal to the ones in the separate analyses, and X^2 values should add up approximately. (There is one slight difference in F(P) due to the replacement of a zero cell). The results are shown in detail below:

	Term	Combined analysis	Vote analysis	Competence analysis
	X^2	18.5384	11.5445	6.5765
	df	18	13	5
	P	0.4207	0.5653	0.2541
1)	MEANC1	49.75	-	49.75
2)	MEANC3	18.41	-	18.41
3)	P1→C1	-31.08	-	-31.08
4)	P1→C3	21.37	-	21.37
5)	P3→C1	24.40	-	24.39
6)	P3→C3	-15.88	-	-15.88
7)	G < P2→C1	8.21	-	8.21
8)	MEAN Reagan	52.51	52.39	-
9)	C1	36.06	36.07	-
10)	P1 < C2	-35.32	-35.71	-
11)	P1 < C13	-11.19	-10.94	-
12)	G < C2	7.34	6.72	-

EXKURS: NONMET - Eingabe und -Ausgabe (Teil B)

Da die Funktionen f_1 und f_2 einerseits und f_3, f_4, f_5 andererseits blockweise unabhängig sind kann die Pfadanalyse auch in zwei getrennten Schritten gerechnet werden mit (P, G, C) bzw. (P, G, C, V) als Ausgangskreuztabellen. Die Anpassung des Gesamtmodells erhält man durch Addition der CHI-Quadrate (Wald-Statistik) und der Freiheitsgrade. Benötigt wird dann lediglich noch eine relativ detaillierte Tabelle der CHI-Quadrat Verteilung zur Bestimmung der Überschreitungswahrscheinlichkeit P. Diese Vorgehensweise dürfte i.a. arbeitsökonomischer sein, da dann weitgehend mit implizierten - also von NONMET intern erzeugten Design-Matrizen - gerechnet werden kann.
Die hier dargestellte Analyse in einem Schritt aus der 6x6 Tafel heraus erfordert im Gegensatz dazu eine explizite Angabe der Design-Matrix. Da eine solche explizite Konstruktion auch in anderem Zusammenhang angezeigt sein kann (etwa bei simultaner Berücksichtigung von auch quantitativen Variablen) wollen wir diesen aufwendigeren Weg hier beispielhaft dokumentieren. Unter Ausnutzung der verschiedenen Matrixeingabe-Optionen bei NONMET läßt sich der Parameterkartenaufwand trotz allem recht gering halten.

```
/.KUENM5 LOGON UA01,Z21,C'WASH',TIME=24
/DO $ZUMA.NONMET,(SPACE=39000)
TITLE    PATH DIAGRAM CPS 80
NP=6,NR=6,A*=6,U*=6,V=12,X=4,K*=13,Y*=5,Q=1/
17 15 3 4     24 27 35 12 4 12 98 110 66 14 2 6 11 10 58 26 5 9 28 16
93 28 1   0 1 1 83 24 2   2 3 5
1,4;2,5;3,6;1;2;3/
1;3;-1,4;-2,5;-3,6/
1 6 11 16 21 26;2 7 12 17 22 27;1 6 -11 -16;2 7 -12 -17;-11 -16 21 26;
-12 -17 22 27;11 -16;
3 4 5 8 9 10 13 14 15 18 19 20 23 24 25 28 29 30;3 -5 8 -10 13 -15 18 -20 23 -25
28 -30;4 9 -24 -29;3 5 8 10 -23 -25 -28 -30;4 -9 24 -29/
990
END
/LOGOFF
```

Die sogenannten Hauptparameter (Karten Nr. 4) lassen sich mit mehr Zeichen, aber dafür systematischer und mnemotechnisch hilfreicher equivalent auch wie folgt schreiben (vgl. NONMET-Manual 6.1 - 6.2.0 und 6.2.1):

NP=6,A*=*NUM,NROWA*=6,RANKX=12,X=NUM,K*=*NUM,NROWK*=5,Q=ID/

Die Matrizen A und K werden also über ihre Diagonalblöcke A* bzw. K* definiert, die zugehörige Zeilenzahl (number of rows) beträgt 6 bzw. 5 und beide Matrizen werden "by number" freiformatig eingegeben. Die Design-Matrix X hat den Rang (=Spaltenzahl bei richtiger Definition) 12 und wird als volle Matrix "by number" definiert. Alle Matrizen werden zeilenweise eingegeben, nur die Designmatrix spaltenweise. Die "by number"-Option ist immer dann günstig, wenn eine Matrix aus vielen Nullen und sonst nur +1 bzw. -1 besteht. Dann braucht man nämlich nur anzugeben, an welcher Stelle eine '1' steht; ein Minuszeichen vor der Spalten- bzw. Zeilenzahl weist einen '-1'-Eintrag aus.

Auf diese Weise lassen sich die Matrizen A und K sehr einfach definieren, lediglich die Matrix X erfordert größeren Aufwand, der aber primär im Aufstellen der Matrix überhaupt liegt. Die volle Design-Matrix ist auf der folgenden Seite wiedergegeben. Zu beachten ist, daß diese Matrix 30 Zeilen hat, da für jede der 6 Subpopulationen 5 Funktionen definiert sind. Die Zeilen 1 - 5 beziehen sich also auf die 1. Subpopulation (gemäß TABLE 5 also "Demokraten", die die Regierungspolitik für "poor" halten), die Zeilen 6 - 10 auf die 2. Subpopulation, etc.
Da die Design-Matrix selbst konstruiert werden muß, werden auch die Effektnamen nicht mehr automatisch geliefert. Sie können aber - was hier unterblieben ist - zur besseren Lesbarkeit des Ausdrucks selbst spezifiziert werden.

USER REQUESTED CONTRASTS:

TESTS OF INDIVIDUAL PARAMETERS:

PARAMETER	B	VARIANCE	CHI SQUARE	P
1	0.4975D+00	0.2383D-03	1038.80	0.00000
2	0.1841D+00	0.1339D-03	253.01	0.00000
3	-0.3108D+00	0.3774D-03	255.99	0.00000
4	0.2137D+00	0.3535D-03	129.25	0.00000
5	0.2440D+00	0.4977D-03	119.64	0.00000
6	-0.1588D+00	0.1673D-03	150.67	0.00000
7	0.8210D-01	0.7652D-03	8.81	0.00300
8	0.5251D+00	0.5730D-04	4811.28	0.00000
9	0.3606D+00	0.3299D-03	394.22	0.00000
10	-0.3532D+00	0.6956D-03	179.36	0.00000
11	-0.1119D+00	0.3018D-03	41.49	-0.00000
12	0.7340D-01	0.6975D-03	7.72	0.00545

****

	7	8	9	10	11	12
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	1.00	1.00	0.00	1.00	0.00
4	0.00	1.00	0.00	1.00	0.00	1.00
5	0.00	1.00	-1.00	0.00	1.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	1.00	1.00	0.00	1.00	0.00
9	0.00	1.00	0.00	1.00	0.00	-1.00
10	0.00	1.00	-1.00	0.00	1.00	0.00
11	1.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	1.00	1.00	0.00	0.00	0.00
14	0.00	1.00	0.00	0.00	0.00	0.00
15	0.00	1.00	-1.00	0.00	0.00	0.00
16	-1.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00
18	0.00	1.00	1.00	0.00	0.00	0.00
19	0.00	1.00	0.00	0.00	0.00	0.00
20	0.00	1.00	-1.00	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00
22	0.00	0.00	0.00	0.00	0.00	0.00
23	0.00	1.00	1.00	0.00	-1.00	0.00
24	0.00	1.00	0.00	-1.00	0.00	1.00
25	0.00	1.00	-1.00	0.00	-1.00	0.00
26	0.00	0.00	0.00	0.00	0.00	0.00
27	0.00	0.00	0.00	0.00	0.00	0.00
28	0.00	1.00	1.00	0.00	-1.00	0.00
29	0.00	1.00	0.00	-1.00	0.00	-1.00
30	0.00	1.00	-1.00	0.00	-1.00	0.00

CHI-SQUARE DUE TO ERROR = 18.5384 DF = 18 P = 0.4207

DESIGN MATRIX (T X V)

	1	2	3	4	5	6
1	1.00	0.00	1.00	0.00	0.00	0.00
2	0.00	1.00	0.00	1.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00
6	1.00	0.00	1.00	0.00	0.00	0.00
7	0.00	1.00	0.00	1.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00
11	1.00	0.00	-1.00	0.00	-1.00	0.00
12	0.00	1.00	0.00	-1.00	0.00	-1.00
13	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00
16	1.00	0.00	-1.00	0.00	-1.00	0.00
17	0.00	1.00	0.00	-1.00	0.00	-1.00
18	0.00	0.00	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00
21	1.00	0.00	0.00	0.00	1.00	0.00
22	0.00	1.00	0.00	0.00	0.00	1.00
23	0.00	0.00	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00	0.00	0.00
26	1.00	0.00	0.00	0.00	1.00	0.00
27	0.00	1.00	0.00	0.00	0.00	1.00
28	0.00	0.00	0.00	0.00	0.00	0.00
29	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	0.00	0.00	0.00	0.00	0.00

Conclusion

Substantively, there is no doubt that party identification is a major factor influencing the evaluation of relative economic competency. Even, if we cannot numerically arrive at figures measuring 'total impact', Miller's (1978) claim receives further warrant . However, due to our analytic approach - circumventing the problem of multicollinearity - we were able to establish the competency evaluation as a major factor directly impacting on the voting decision.

In respect to data analytic techniques, we feel that the GSK approach offers a wide range of different conceptualizations[21]; and especially in employing linear models it produces coefficients that are easy to grasp in substantive terms. However, there are competing philosophies in data analysis. People more interested in strict model testing rather than analytical description might be well advised using maximum-likelihood techniques instead.[22]

Notes

+ The data utilized in this analysis were provided by the Inter-University Consortium for Political and Social Research and collected by the University of Michigan's Center for Political Studies. Neither the providers nor the collectors bear any responsibility for our analysis.

1) An overview on competing findings is given in Wides and Kuechler (1981).

2) A detailed exposition of these findings can be found in Wides (1979), Kuechler and Wides (1980) as well as Wides and Kuechler (1981). However, the presentation in this paper is based on a reanalysis of the '76 data to improve comparability with the '80 analysis.

3) Out of 1614 cases in the pre-election wave 1408 respondents could be successfully reinterviewed. 877 out of 962 voting at all cast their ballots in favor of the two major candidates. Our analysis is based on this subsample.

4) All variable numbers refer to CPS's codebook information on the data file.

5) Interesting in itself, but not pursued in this paper, the pre- and post-measures of party identification show considerable instability. We relied on the pre-election measure.

6) In the marginal distribution about one half of the respondents choose category "4" which was explicitly labelled "fair job". In general, the seeming loss of detail in grouping variables is well warranted given that the metric nature of these scales is highly dubious.

7) We do not intend to describe this approach in detail. The interested reader is referred to the literature cited above.

8) For more details on this relationship see Kuechler (1980).

9) Readers more interested in applying this data analytic approach will find a persuasive introduction in a recently published textbook (Forthofer/Lehnen 1981). Also, the extensive bibliography renders an all but complete review of the literature on the subject, both formal and applied.

10) An example for a more complex function can be found in a later section of this paper.

11) For practical computations two computer programs are available. We are using Herbert Kritzer's NONMET, which is more suited to the user than Richard Landis' GENCAT (MISCAT). Also, in SAS the procedure FUNCAT is available.

12) The process of variable selection in non-metric multivariate analysis is described in Higgins and Koch (1977).

13) Alternatively, a 1/0 coding could be employed. Of course, this would not effect the goodness-of-fit, but rather the size of the single coefficients which would receive a different substantive interpretation.

14) It should be pointed out that our analysis is restricted to actual voters. Thus our findings would not necessarily contradict any claims of weakening party affiliation in general.

15) Of course the covariance matrix of P and consequently of F(P) uses this kind of information. Roughly speaking, proportions from subpopulations with less cases will be allowed greater residuals when estimating a particular model.

16) Thus, in our case, we could have looked at competency effects contingent upon level of party identification simultaneously claiming an overall party effect. This, of course, is not a matter of statistical but of substantive considerations.

17) More detailed advice on this is given by Forthofer and Lehnen (1981).

18) The procedures developed by Higgins and Koch (1977) can serve as a very useful screening device in earlier steps of the analysis.

19) It seems that this option has not been extensively explored so far. In earlier publications (e.g. Kuechler 1979) we claimed the handling of "path diagrams" to be a distinct advantage of the Goodman approach over the GSK approach.

20) The separate analysis on competency as dependent is not reported in detail here.

21) For instance, to be on the safe side with the partly rather heavily skewed subpopulation distributions we ran parallel log-odd analyses that reconfirmed our model.

22) Besides techniques in the "Goodman" tradition, the approach developed by Nelder and Wedderburn (1972) in Great Britain deserves special attention. Unfortunately, the corresponding computer program GLIM (Baker and Nelder 1978) is hard to use for non-statisticians.

References

Baker, R.J. and J.A. Nelder (1978). The GLIM system. London: Numerical Algorithm Comp.

Brody, Richard A., and Paul M. Sniderman (1977). From Life Space to Polling Place: The Relevance of Personal Concerns for Voting Behavior. British Journal of Political Science 7:337-60.

Duncan, Otis D. (1975). Introduction to Structural Equation Models. New York: Academic Press.

Fienberg, Stephen E. (1977). The Analysis of Cross-Classified Categorical Data. Cambridge, Mass.: MIT Press.

Fiorina, Morris P. (1978). Economic Retrospective Voting in American National Elections: A Micro-Analysis. American Journal of Political Science 22:426-43.

Forthofer, Ron N., and Robert G. Lehnen (1981). Public Program Analysis: A New Categorical Data Approach. Belmont, Cal.: Wadsworth.

Goodman, Leo A., and J. Magidson (1978). Analyzing Qualitative/ Categorical Data: Log-linear Models and Latent Structure Analysis. Cambridge, Mass.: ABT Books.

Grizzle, James E., C. Frank Starmer, and Gary G. Koch (1969). Analysis of Categorical Data by Linear Models. Biometrics 25:489-504.

Higgins, James E., and Gary G. Koch (1977). Variable Selection and Generalized Chi-Square Analysis of Categorical Data applied to a Large Cross-Sectional Occupational Health Study. International Statistical Review 45:51-62.

Kinder, Donald R. and D. Roderick Kiewiet (1979). Economic Discontent and Political Behavior: The Role of Personal Grievances and Collective Economic Judgments in Congressional Voting. American Journal of Political Science 23:495-527.

Kuechler, Manfred (1979). Multivariate Analyseverfahren. Stuttgart: B.G. Teubner.

Kuechler, Manfred (1980). The Analysis of Non-Metric Data: The Relation to Dummy Dependent Variable Regression using an Additive Saturated Grizzle-Starmer-Koch Model. Sociological Methods & Research 8:369-388.

Kuechler, Manfred, and Jeffrey W. Wides (1980). Perzeption der Wirtschaftslage und Wahlentscheidung. Politische Vierteljahresschrift 21:4-19.

Miller, Arthur H. (1978). Partisanship Reinstated? A Comparison of the 1972 and 1976 U.S. Presidential Elections. British Journal of Political Science 8:129-52.

Nelder, J.A., and R.W.M. Wedderburn (1972). Generalized Linear Models. Journal Royal Statistical Society A 135:370-384.

Reynolds, H.T. (1977a). The Analysis of Cross-Classifications. New York: Free Press.

Swafford, Michael (1980). Parametric Techniques for Contingency Table Analysis. American Sociological Review 45:664-690.

Wahlke, John C. (1979). Pre-Behavioralism in Political Science. American Political Science Review 73:9-31.

Wides, Jeffrey W. (1979). Perceived Economic Competency and the Ford/Carter Election. Public Opinion Quarterly 43: 548-556.

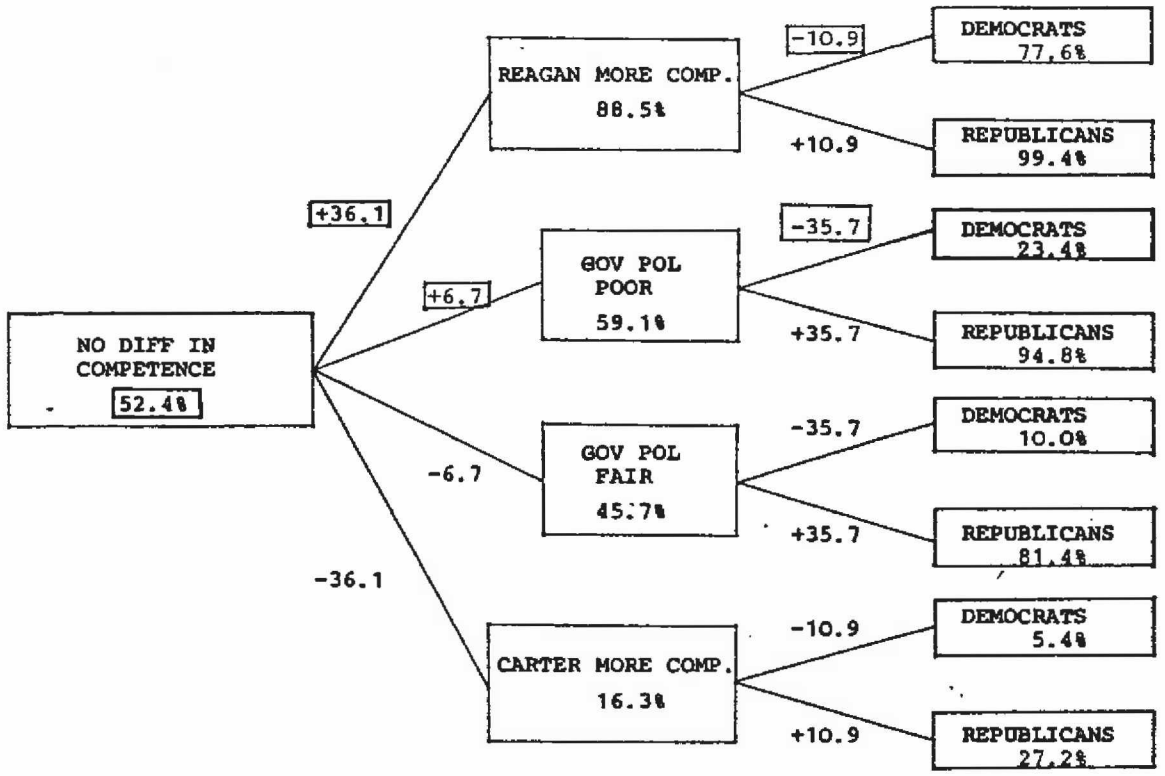
Wides, Jeffrey W., and Manfred Kuechler (1981). Economic Perceptions and the '76 Presidential Vote: A WLS Analysis. Micropolitics (forthcoming).

TEILNEHMERLISTE

Name	Institution
1) Hans Jürgen Andress	Fak. f. Soz. Universität Bielefeld
2) Hans Peter Bäumer	Rechenzentrum, Uni Oldenburg
3) Helmut Bender	EWH, Rheinland-Pfalz, Landau
4) Wolfgang Blass-Wilhelms	MPI, Freiburg
5) Michael Braun	München
6) Frau Burow-Auffarth	Inst.f.Soz., Uni Hamburg
7) H. Busse	Bundesgesundheitsamt, Berlin
8) Robert Danziger	Universität Mannheim
9) Klaus Echterhagen	GHS Wuppertal
10) Frank Faulbaum	ZUMA
11) Ulrich Fischer	Soz.Inst. Erlangen-Nürnberg
12) Friedrich Förster	Konsumenteninformation, Uni Mannheim
13) Gerhard Franz	Universität Mannheim
14) Johann Haefele	Städtebau-Institut Stuttgart
15) Bernhard Hesener	KFN Hannover
16) Ursula Hoffmann-Lange	Universität Mannheim
17) Gerhard Hofmann	FB 3, Universität Frankfurt
18) Alexander Hoschka	Inst.f.Soz. Uni München
19) Thomas Kohlmann	Med.Soz. Uni Marburg
20) Udo Kuckartz	FU Berlin
21) Helga Leitner	Inst.f. Geographie Uni Wien
22) Günter Marxen	Rechenzentrum Köln
23) Herbert Matschinger	Med.Soz. Universität Marburg
24) Heiner Meulemann	Zentralarchiv Köln
25) Gerhard Paaß	GMD St. Augustin
26) Dieter Rondorf	Niederkassel
27) Silke Schmidt	Seminar f. Soz. Uni Hamburg
28) Volkhart Schönberg	MPI, Freiburg
29) Robert Schreieck	Mannheim
30) Erhard Schwedler	FB 2 Universität Frankfurt
31) Irmfried Speiser	Inst.f.Soz. Universität Wien
32) Joachim Werner	Psych.Inst. Uni Heidelberg
33) Helmut Wohlschlägl	Inst.f. Geographie, Universität Wien

BLE 2:

APHICAL PRESENTATION OF BEST MODEL FOR RERCENTAGE REAGAN VOTE '80



χ^2 = 11.54
DF = 13 (out of 18)
P = .5653

156

TABLE 3: COMPARISON OF '76 AND '80 RESULTS

Effects	1976	1980	
MEAN	54.7%	52.4%	
P1 < C13	-10.2%	-10.9%	
P1 < C2	-25.8%	-35.7%	
C1	33.3%	36.1%	(C1)
C3	-39.3%		
G < C3	2.3%	6.7%	(G < C2)
x^2	14.82	11.54	
df	12	13	
P	.2516	.5653	

All effects significant on at least 1%-level, except G < C3 (1976) on 5%-level only.

In both 1976 and 1980 the Republican candidate's voting percentage is analyzed, in order to show the effect of G on the incumbent, the 1976 order of G categories has been reversed for 1980.